第28卷 增刊1　　工 程 数 学 学 报　　Vol. 28 Supp.1
2011年11月　　CHINESE JOURNAL OF ENGINEERING MATHEMATICS　　Nov. 2011

文章编号:1005-3085(2011)07-0001-42

2011高教社杯全国大学生数学建模竞赛

谢金星

(清华大学，北京 100084)

2011年是全国大学生数学建模竞赛20周年．这项每年一次的竞赛由中国工业与应用数学学会于1992年创办，从1994年起由教育部高教司和中国工业与应用数学学会共同主办，2002年起由高等教育出版社独家冠名赞助．在教育部领导“扩大受益面，保证公正性，推动教育改革”的指示精神指导下，在各级教育行政部门和广大教师的积极指导和参与下，二十年来这项竞赛发展迅速．20年前开始举办这项竞赛时，只有74所高校的314队参加；到今年，参赛学校数达到1251所，超过了我国现有高校总数的一半；参赛队数达19490队(参赛学生58000多人)，参赛学校数和队数分别每年平均以超过16%及24%的速度增长．数学建模竞赛不仅已成为我国高校历史最悠久、规模最大、影响广泛的学科性竞赛活动，而且成功地推动了数学教学的改革，造就了这些年来规模最大也最成功的、以推进数学建模为核心的数学教学改革实践．

2011高教社杯全国大学生数学建模竞赛(即第20届竞赛)于2011年9月9日至12日举行．来自全国33个省/市/自治区(包括香港和澳门特区)以及新加坡和美国的1251所高校19490个队(其中本科组16008队、专科组3248队)的近六万多名大学生参加了今年的比赛，是历年来参赛人数最多的，与2010年的1196所院校的17317队(其中本科组14119队、专科组3198队)相比，校数增长4.6%，队数增长12.5%(本专科队数分别增长13.4%和8.9%)．竞赛继续得到了北京迈斯沃克软件有限公司的赞助与支持，并设立了Matlab创新奖．

竞赛答卷首先在由省/市/自治区形成的27个赛区和联合赛区(今年由内蒙、宁夏、青海、西藏组成)进行初评，评出各赛区的获奖者，然后各赛区按一定比例将优秀答卷送全国竞赛组委会(与去年一样，对境外参赛者来说，本项竞赛也被称为“当代大学生数学建模竞赛”，论文单独进行评奖)．送全国评阅的论文总数是1821份(A, B, C, D题分别为1037, 463, 204, 117份)，约占报名参赛队数的9%，A题还有1篇按创新点论文送评．全国组委会聘请了48位专家组成全国评阅专家组，认真评阅了候选论文，也对一些队进行了面试答辩．经过全国评阅专家的评阅与全国组委会的复核、面试，以及异议期的审核，按照从严控制获奖比例特别是全国一等奖数量的原则，确定获全国奖共1536队，获奖比例约为8%，与去年相比基本持平．其中本科组一等奖224队，二等奖1040队，专科组一等奖51队，二等奖221队，一、二等奖分别占参赛总数的1.5%和6.5%．高教社杯的获得者本科组为南京信息工程大学黄思、盛振峰、周敏强，专科组为江西应用技术职业学院杨忠、张岐良、徐小辉．Matlab创新奖获得者本科组为第三军医大学唐棣、董小小、魏歆，专科组为九江学院乔衡山、许琼、徐景华．同时，北京、天津、山西、上海、江苏、浙江、山东、广东、重庆、四川十个赛区组委会获得优秀组织工作奖．

这项竞赛之所以受到大学生们如此热烈的欢迎，是因为它有以下的特点：赛题由工程技术、管理科学、社会生活等领域的实际问题简化加工而成，要求参赛者结合实际问题灵活运用数学和计算机软件以及其他学科的知识，通过建立、求解、评估、改善数学模型，充分发挥其

聪明才智和创造精神；三名大学生组成一队，团结合作，选择一题在三天时间内完成一篇研究论文；可以自由地收集、查阅资料，调查研究，使用计算机、互联网和各种软件(但是不能与队外的任何人讨论赛题)；赛题没有事先确定的答案，论文评阅的标准是：假设的合理性、建模的创造性、结果的正确性和表述的清晰性．

为了进一步加强竞赛规则的执行，促进竞赛的健康发展，要求参赛同学签名承诺："我们完全明白，在竞赛开始后参赛队员不能以任何方式(包括电话、电子邮件、网上咨询等)与队外的任何人(包括指导教师)研究、讨论与赛题有关的问题．我们郑重承诺，严格遵守竞赛纪律，以保证竞赛的公平、公正性．如有违反竞赛规则的行为，我们将会受到严肃处理"．

今年A题"城市表层土壤重金属污染分析"由山东理工大学李功胜教授提出，B题"交巡警服务平台的设置与调度"由解放军后勤工程学院但琦教授和解放军信息工程大学韩中庚教授提出，C题"企业退休职工养老金制度的改革"由济南大学许振宇教授提出，D题"天然肠衣搭配问题"由复旦大学陆立强教授提出，在此一并表示感谢．为了更广泛、有效地收集适合竞赛的题目和素材，全国大学生数学建模竞赛组委会再次向全社会诚征赛题，有意者请与全国组委会副秘书长蔡志杰联系，将赛题或素材用电子邮件直接发给蔡志杰教授(Email: zhijiecai@163.com)，并请在邮件主题(Subject)中注明"赛题素材"，在邮件正文中写明单位、姓名、电话等信息，以便联系．同时，务必注意保密，切勿将与素材有关的任何内容与其他人交流．

为了与广大师生进行交流，对今后的竞赛予以适当引导，全国评阅专家组选择了一些优秀论文在本刊发表(包括新加坡国立大学的一篇参加"当代大学生数学建模竞赛"的获奖论文)，并请命题者或评阅者撰文讲评．

发表的论文是同学们三天内写出的，为了保持原貌只作了适当的删节和文字上的修正，文章不可避免地存在着相当多的不妥之处，请读者谅解．

希望对这项竞赛作进一步了解的读者，请访问竞赛网站：http://mcm.edu.cn.

A题 城市表层土壤重金属污染分析

随着城市经济的快速发展和城市人口的不断增加，人类活动对城市环境质量的影响日显突出．对城市土壤地质环境异常的查证，以及如何应用查证获得的海量数据资料开展城市环境质量评价，研究人类活动影响下城市地质环境的演变模式，日益成为人们关注的焦点．

按照功能划分，城区一般可分为生活区、工业区、山区、主干道路区及公园绿地区等，分别记为1类区、2类区、···、5类区，不同的区域环境受人类活动影响的程度不同．

现对某城市城区土壤地质环境进行调查．为此，将所考察的城区划分为间距1公里左右的网格子区域，按照每平方公里1个采样点对表层土(0–10厘米深度)进行取样、编号，并用GPS记录采样点的位置．应用专门仪器测试分析，获得了每个样本所含的多种化学元素的浓度数据．另一方面，按照2公里的间距在那些远离人群及工业活动的自然区取样，将其作为该城区表层土壤中元素的背景值．

附件1列出了采样点的位置、海拔高度及其所属功能区等信息，附件2列出了8种主要重金属元素在采样点处的浓度，附件3列出了8种主要重金属元素的背景值．

现要求你们通过数学建模来完成以下任务：

1) 给出8种主要重金属元素在该城区的空间分布，并分析该城区内不同区域重金属的污染程度；

2) 通过数据分析，说明重金属污染的主要原因；

3)　分析重金属污染物的传播特征，由此建立模型，确定污染源的位置；

4)　分析你所建立模型的优缺点，为更好地研究城市地质环境的演变模式，还应收集什么信息？有了这些信息，如何建立模型解决问题？

注：附件数据略，请从 http:// mcm.edu.cn 下载.

B题　交巡警服务平台的设置与调度

“有困难找警察”，是家喻户晓的一句流行语．警察肩负着刑事执法、治安管理、交通管理、服务群众四大职能．为了更有效地贯彻实施这些职能，需要在市区的一些交通要道和重要部位设置交巡警服务平台．每个交巡警服务平台的职能和警力配备基本相同．由于警务资源是有限的，如何根据城市的实际情况与需求合理地设置交巡警服务平台、分配各平台的管辖范围、调度警务资源是警务部门面临的一个实际课题．

试就某市设置交巡警服务平台的相关情况，建立数学模型分析研究下面的问题：

1)　附件1中的附图1给出了该市中心城区A的交通网络和现有的20个交巡警服务平台的设置情况示意图，相关的数据信息见附件2．请为各交巡警服务平台分配管辖范围，使其在所管辖的范围内出现突发事件时，尽量能在3分钟内有交巡警(警车的时速为60 km/h)到达事发地点．

对于重大突发事件，需要调度全区20个交巡警服务平台的警力资源，对进出该区的13条交通要道实现快速全封锁．实际中一个平台的警力最多封锁一个路口，请给出该区交巡警服务平台警力合理的调度方案．

根据现有交巡警服务平台的工作量不均衡和有些地方出警时间过长的实际情况，拟在该区内再增加2至5个平台，请确定需要增加平台的具体个数和位置．

2)　针对全市(主城六区A, B, C, D, E, F)的具体情况，按照设置交巡警服务平台的原则和任务，分析研究该市现有交巡警服务平台设置方案(参见附件)的合理性．如果有明显不合理，请给出解决方案．

如果该市地点P(第32个节点)处发生了重大刑事案件，在案发3分钟后接到报警，犯罪嫌疑人已驾车逃跑．为了快速搜捕嫌疑犯，请给出调度全市交巡警服务平台警力资源的最佳围堵方案．

附件1：A区和全市六区交通网络与平台设置的示意图．

附件2：全市六区交通网络与平台设置的相关数据表(共5个工作表)．

注：附件数据略，请从 http://mcm.edu.cn 下载．

C题　企业退休职工养老金制度的改革

养老金也称退休金，是一种根据劳动者对社会所作贡献及其所具备享受养老保险的资格，以货币形式支付的保险待遇，用于保障职工退休后的基本生活需要．

我国企业职工基本养老保险实行“社会统筹”与“个人账户”相结合的模式，即企业把职工工资总额按一定比例(20%)缴纳到社会统筹基金账户，再把职工个人工资按一定比例(8%)缴纳到个人账户．这两个账户我们合称为养老保险基金．退休后，按职工在职期间每月(或年)的缴费工资与社会平均工资之比(缴费指数)，再考虑到退休前一年的社会平均工资等因素，从社会统筹账户中拨出资金(基础养老金)，加上个人工资账户中一定比例的资金(个人账户养老金)，作为退休后每个月的养老金．养老金会随着社会平均工资的调整而调整．如果

职工死亡，社会统筹账户中的资金不退给职工，个人账户中的余额可继承．个人账户储存额以银行当时公布的一年期存款利率计息，为简单起见，利率统一设定为3%．

养老金的发放与职工在职时的工资及社会平均工资有着密切关系；工资的增长又与经济增长相关．近30年来我国经济发展迅速，工资增长率也较高；而发达国家的经济和工资增长率都较低．我国经济发展的战略目标，是要在21世纪中叶使我国人均国民生产总值达到中等发达国家水平．

现在我国养老保险改革正处于过渡期．养老保险管理的一个重要的目标是养老保险基金的收支平衡，它关系到社会稳定和老龄化社会的顺利过渡．影响养老保险基金收支平衡的一个重要因素是替代率．替代率是指职工刚退休时的养老金占退休前工资的比例．按照国家对基本养老保险制度的总体思路，未来基本养老保险的目标替代率确定为58.5%．替代率较低，退休职工的生活水准低，养老保险基金收支平衡容易维持；替代率较高，退休职工的生活水准就高，养老保险基金收支平衡较难维持，可能出现缺口．所谓缺口，是指当养老保险基金入不敷出时出现的收支之差．

附件1是山东省职工历年平均工资数据；附件2是2009年山东省某企业各年龄段职工的工资分布情况，附件3是养老金的计算办法．请建立数学模型，解决如下问题：

问题一：对未来中国经济发展和工资增长的形势做出你认为是简化、合理的假设，并参考附件1，预测从2011年至2035年的山东省职工的年平均工资．

问题二：根据附件2计算2009年该企业各年龄段职工工资与该企业平均工资之比．如果把这些比值看作职工缴费指数的参考值，考虑该企业职工自2000年起分别从30岁、40岁开始缴养老保险，一直缴费到退休(55岁,60岁,65岁)，计算各种情况下的养老金替代率．

问题三：假设该企业某职工自2000年起从30岁开始缴养老保险，一直缴费到退休(55岁,60岁,65岁)，并从退休后一直领取养老金至75岁死亡．计算养老保险基金的缺口情况，并计算该职工领取养老金到多少岁时，其缴存的养老保险基金与其领取的养老金之间达到收支平衡．

问题四：如果既要达到目标替代率，又要维持养老保险基金收支平衡，你认为可以采取什么措施．请给出你的理由．

注：附件数据略，请从http://mcm.edu.cn下载．

D题　天然肠衣搭配问题

天然肠衣(以下简称肠衣)制作加工是我国的一个传统产业，出口量占世界首位．肠衣经过清洗整理后被分割成长度不等的小段(原料)，进入组装工序．传统的生产方式依靠人工，边丈量原料长度边心算，将原材料按指定根数和总长度组装出成品(捆)．

原料按长度分档，通常以0.5米为一档，如：3米–3.4米按3米计算，3.5米–3.9米按3.5米计算，其余的依此类推．表1是几种常见成品的规格，长度单位为米，∞表示没有上限，但实际长度小于26米．

为了提高生产效率，公司计划改变组装工艺，先丈量所有原料建立一个原料表．表2为某批次原料描述．根据成品和原料描述，设计一个原料搭配方案，工人根据这个方案"照方抓药"进行生产．

公司对搭配方案有以下具体要求：

1)　对于给定的一批原料，装出的成品捆数越多越好；
2)　对于成品捆数相同的方案，最短长度最长的成品越多，方案越好；
3)　为提高原料使用率，总长度允许有±0.5米的误差，总根数允许比标准少1根；

4)　某种规格对应原料如果出现剩余，可以降级使用．如长度为 14 米的原料可以和长度介于 7 米–13.5 米的进行捆扎，成品属于 7 米–13.5 米的规格；

5)　为了食品保鲜，要求在 30 分钟内产生方案．

请建立上述问题的数学模型，给出求解方法，并对表 1 和表 2 给出的实际数据进行求解，给出搭配方案．

表 1:　成品规格表

最短长度 (米)	最大长度 (米)	根数	总长度 (米)
3	6.5	20	89
7	13.5	8	89
14	∞	5	89

表 2:　原料描述表

长度 (米)	3–3.4	3.5–3.9	4–4.4	4.5–4.9	5–5.4	5.5–5.9	6–6.4	6.5–6.9
根数	43	59	39	41	27	28	34	21
长度 (米)	7–7.4	7.5–7.9	8–8.4	8.5–8.9	9–9.4	9.5–9.9	10–10.4	10.5–10.9
根数	24	24	20	25	21	23	21	18
长度 (米)	11–11.4	11.5–11.9	12–12.4	12.5–12.9	13–13.4	13.5–13.9	14–14.4	14.5–14.9
根数	31	23	22	59	18	25	35	29
长度 (米)	15–15.4	15.5–15.9	16–16.4	16.5–16.9	17–17.4	17.5–19.9	18–18.4	18.5–18.9
根数	30	42	28	42	45	49	50	64
长度 (米)	19–19.4	19.5–19.9	20–20.4	20.5–20.9	21–21.4	21.5–21.9	22–22.4	22.5–22.9
根数	52	63	49	35	27	16	12	2
长度 (米)	23–23.4	23.5–23.9	24–24.4	24.5–24.9	25–25.4	25.5–25.9		
根数	0	6	0	0	0	1		

2011 高教社杯全国大学生数学建模竞赛获奖名单

高教社杯获得者：　本科组：黄　思、盛振峰、周敏强 (南京信息工程大学)

　　　　　　　　　专科组：杨　忠、张岐良、徐小辉 (江西应用技术职业学院)

Matlab 创新奖获得者：　本科组：唐　棣、董小小、魏　歆 (第三军医大学)

　　　　　　　　　　　专科组：乔衡山、许　琼、徐景华 (九江学院)

第28卷 增刊1
2011年11月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 28 Supp.1
Nov. 2011

文章编号:1005-3085(2011)07-0043-07

表层土壤中重金属污染分析

蔡志杰[1,2]

(1- 复旦大学数学科学学院，上海 200433；2- 上海市现代应用数学重点实验室，上海 200433)

摘 要: 本文对2011年全国大学生数学建模竞赛A题进行分析，给出各问题的求解方法，并对学生的论文进行评述.

关键词: 重金属污染；散乱数据插值；评价指标；对流方程；特征线法

分类号: AMS(2000) 65D05; 62H12; 80A23; 62P12 **中图分类号**: O241.3; O212.4; X8 **文献标识码**: A

1 问题的提出

2011年全国大学生数学建模竞赛A题“表层土壤中重金属污染分析”来自于环境监测部门的一个实际问题．题目给出了一组采样点上8种重金属元素的浓度值，要求作出污染程度的评价、分析污染原因、确定污染源的位置．本文给出各问题的求解方法，并对评阅中发现的问题进行分析与评述.

2 地形高度函数和浓度分布函数的插值方法

首先观察各采样点的分布情况，如图1所示，其中“•”为生活区，“♦”为工业区，“■”为山区，“▼”为交通区，“★”为公园绿地区．各采样点不是分布在规则的网格点上，而是呈现出散乱状态，因而应采用散乱数据插值.

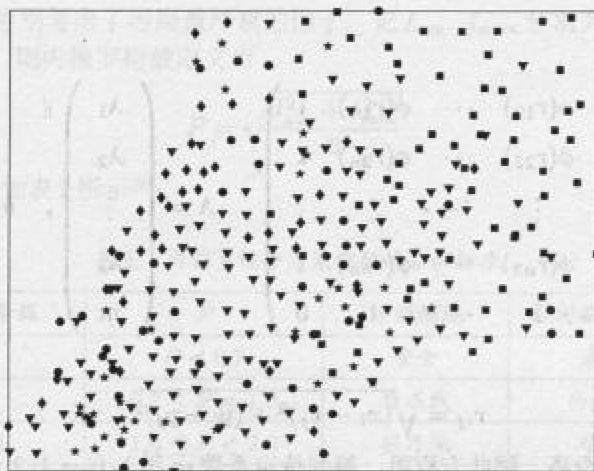


图1: 采样点分布图

一种插值方法是Shepard插值[1,2]，其基本思想是：某一点的函数值受周围各点的影响，较近的点影响较大，较远的点影响较小，其影响权数与距离平方成反比，因而又称为距离平方反比律．具体方法如下：

记采样点为 $P_i(x_i,y_i)$，其高度为 $h_i\,(i=1,2,\cdots,n)$，其中 n 为采样点数．设地形高度函数为 $h(x,y)$，则对任一点 $P(x,y)$，其高度为

$$h(x,y)=\begin{cases} h_j, & \text{当某个 } r_j=0 \text{ 时}, \\ \sum\limits_{i=1}^{n}\frac{h_i}{r_i^2}\Big/\sum\limits_{i=1}^{n}\frac{1}{r_i^2}, & \text{其它}, \end{cases} \tag{1}$$

其中

$$r_i=\sqrt{(x-x_i)^2+(y-y_i)^2} \tag{2}$$

为 P 点与 P_i 点之间的距离．

另一种常用的散乱数据插值方法是Kriging插值[2]，它建立在地统计学理论基础上，利用区域化变量的原始数据和半方差函数的结构特征，对位采样点的区域化变量的取值进行线性最佳无偏估计．

设任一点 $P(x,y)$ 的高度为

$$h(x,y)=m+\sum_{i=1}^{n}\lambda_i\phi(r_i), \tag{3}$$

其中 P 点与 P_i 点之间的距离 r_i 由 (2) 式给出，m 和 $\lambda_i\,(i=1,2,\cdots,n)$ 为待定系数，满足

$$\sum_{i=1}^{n}\lambda_i=1 \tag{4}$$

及插值条件

$$h(x_i,y_i)=h_i,\quad i=1,2,\cdots,n. \tag{5}$$

由此得到关于 m 和 $\lambda_i\,(i=1,2,\cdots,n)$ 的线性代数方程组

$$A\Lambda=b, \tag{6}$$

其中

$$A=\begin{pmatrix} \phi(r_{11}) & \phi(r_{12}) & \cdots & \phi(r_{1n}) & 1 \\ \phi(r_{21}) & \phi(r_{22}) & \cdots & \phi(r_{2n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi(r_{n1}) & \phi(r_{n2}) & \cdots & \phi(r_{nn}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix},\quad \Lambda=\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ m \end{pmatrix},\quad b=\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \\ 1 \end{pmatrix}, \tag{7}$$

这里

$$r_{ij}=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2} \tag{8}$$

为 P_i 点与 P_j 点之间的距离．解此方程组，得到待定系数 m 和 $\lambda_i\,(i=1,2,\cdots,n)$ 的值，然后利用插值函数 (3) 即可得到高度函数 $h(x,y)$．

用同样的方法对重金属元素的浓度进行插值，可以得到各重金属元素的浓度分布函数．

3　污染程度评价指标

土壤重金属污染评价的方法很多，如地质累积指数[3,4]、内梅罗指数法[5]等.

3.1　地质累积指数

地质累积指数，又称为Muller指数，是Muller于1949年提出的针对沉积物种的重金属污染程度的评价方法．它不仅考虑了自然地质过程造成的背景值影响，而且也充分注意了人为活动对重金属污染的影响，因此，这一指数不仅反映了重金属分布的自然变化特征，而且可以判别人为活动对环境的影响，是区分人为活动影响的重要参数[3,4]．地质累积指数定义为

$$I_{\mathrm{geo}} = \log_2 \frac{c}{k\mu}, \tag{9}$$

其中c为重金属元素浓度的实际测量值，μ为重金属元素的背景值，k为修正指数，通常取为1.5.

不同的地质累积指数I_{geo}对应于不同的污染程度，其分级标准参见表1[3].

表1:　地质累积指数分级表

地质累积指数	地质累积指数分级	污染程度
$I_{\mathrm{geo}} < 0$	0	无污染
$0 \le I_{\mathrm{geo}} < 1$	1	无污染–中度污染
$1 \le I_{\mathrm{geo}} < 2$	2	中度污染
$2 \le I_{\mathrm{geo}} < 3$	3	中度污染–强污染
$3 \le I_{\mathrm{geo}} < 4$	4	强污染
$4 \le I_{\mathrm{geo}} < 5$	5	强污染–极强污染
$I_{\mathrm{geo}} \ge 5$	6	极强污染

3.2　内梅罗指数

为了评价各功能区重金属污染的程度，需要将各单个指标进行合成，即构造综合指标．常用的环境质量综合评价指标是内梅罗指数[5]，这是一种兼顾极值或称突出最大值的计权型多因子环境质量指数，即特别考虑了污染最严重的因子．记I_{ave}，I_{max}分别为各单因子环境质量指数的平均值和最大值，则内梅罗指数定义为

$$P = \sqrt{\frac{I_{\mathrm{ave}}^2 + I_{\mathrm{max}}^2}{2}}. \tag{10}$$

污染指数分级方法如表2所示[6]:

表2:　内梅罗综合污染指数的分级标准

污染等级	P	污染描述	污染程度
I	$P < 0.7$	安全	清洁
II	$0.7 \le P < 1.0$	警戒线	尚清洁
III	$1.0 \le P < 2.0$	轻污染	超标
IV	$2.0 \le P < 3.0$	中污染	中度污染
V	$P \ge 3.0$	重污染	严重污染

4 污染原因分析

4.1 相关性分析

为了分析污染原因，考虑不同重金属元素之间是否存在一定的联系．如果不同的重金属存在很高的相关性，那么它们可能存在共生性，因而有理由相信，它们是由同一污染源产生的，从而可以分析出污染产生的原因．

相关系数定义为

$$\rho=\frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2}\sqrt{\sum(y-\overline{y})^2}}. \tag{11}$$

4.2 聚类分析

相关系数仅考虑重金属元素两两之间的关系，要研究多种重金属元素之间的关系，可采用聚类分析[7]的方法．

聚类分析方法是基于数值分类法的思想建立起来的，又称为系统聚类法．它是研究样品或指标分类的一种多元统计分析方法．这里是基于样本指标进行聚类，称为R型聚类，其步骤为：

步骤1 将各重金属元素视为不同的n类，每一类包含一个元素；

步骤2 计算各指标两两之间的相关性$\rho(x_i,x_j)\,(i,j=1,\cdots,n)$；

步骤3 将相关性最大的两类合并成一类；

步骤4 计算新合成的类与其他各类之间的相关系数；

步骤5 重复步骤3和步骤4，直到将元素合并成我们所需要的类．

两类之间的相关系数可以用广义相关系数[8]来计算．设$a_1,a_2,\cdots,a_m$和$b_1,b_2,\cdots,b_n$为给定的两组p维列向量，记

$$x=(a_1,a_2,\cdots,a_m)_{p\times m},\quad y=(b_1,b_2,\cdots,b_n)_{p\times n}. \tag{12}$$

令

$$M_{xx}=E(x-Ex)^{\top}(x-Ex),\quad M_{yy}=E(y-Ey)^{\top}(y-Ey), \tag{13}$$

$$M_{yx}^{\top}=M_{xy}=E(x-Ex)^{\top}(y-Ey). \tag{14}$$

用$\lambda_1,\lambda_2,\cdots,\lambda_l$表示$M_{xx}^{+}M_{xy}M_{yy}^{+}M_{yx}$的全部非零特征根，其中$A^+$表示矩阵$A$的广义逆．定义$x$与$y$之间的广义相关系数为

$$\rho^*(x,y)=\frac{1}{l}\sum_{i=1}^{l}\lambda_i. \tag{15}$$

可以证明：当$m=n=1$时，广义相关系数即为相关系数的平方[8]，即$\rho^*=\rho^2$．

5 污染源位置的确定

为了确定污染源位置，首先需要分析污染物传播的特征．由于我们主要考虑表层土壤，地表径流冲刷为主要传播途径，因而对流是最主要传播特征，而扩散所起的作用相对来说较小．

记$u=u(x,y,t)$为t时刻位于$P(x,y)$点处重金属元素的浓度，利用质量守恒定律，可以得到对流方程[9]

$$\frac{\partial u}{\partial t}+\nabla\cdot\mathbf{q}=0. \tag{16}$$

由于重金属元素的传播主要是由于雨水冲刷引起的，而雨水只能从高处往低处流动，因而可假设流量 $\mathbf{q}$ 与地形海拔高度 h 的梯度成比例，即 $\mathbf{q} \propto \nabla h$.

此外，假设流量 $\mathbf{q}$ 与污染元素浓度成正比应是合理的，即 $\mathbf{q} \propto u$，因此，可令

$$\mathbf{q} = -ku\nabla h, \tag{17}$$

这里负号表示污染元素从高处随雨水向低处传播，$k\,(>0)$ 为比例系数.

将 (17) 式代入 (16) 式，得到

$$\frac{\partial u}{\partial t} - k\nabla h \cdot \nabla u - ku\nabla^2 h = 0. \tag{18}$$

假设时刻 T 测得的各观测点处污染元素的浓度为 $u_i\,(i = 1, 2, \cdots, n)$，则初始条件可写为

$$t = T : u(x_i, y_i) = u_i, \quad i = 1, 2, \cdots, n. \tag{19}$$

由此得到了重金属污染的传播模型 (18)–(19)，模型的具体推导过程可参见文献 [10]. 这是一个一阶双曲型方程，可以通过特征线法求解.

由于只有一个时间点上的截面数据，无法确定系数 k 的值，因此对时间 t 作伸缩变换，即令 $t' = kt$，则方程 (18) 可改写为 (为方便起见，仍记 t' 为 t)

$$\frac{\partial u}{\partial t} - \nabla h \cdot \nabla u - u\nabla^2 h = 0. \tag{20}$$

与 (18) 相比，实际上是令 $k = 1$.

公式 (20) 的特征线满足

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = -\nabla h, \tag{21}$$

若给定初始条件

$$t = T : \mathbf{x} = \mathbf{x}_T, \tag{22}$$

则记过给定点 $(T, \mathbf{x}_T)$ 的特征线为 $\mathbf{x}(t; \mathbf{x}_T)$. 沿着特征线，重金属元素浓度 u 满足

$$\frac{\mathrm{d}u}{\mathrm{d}t} = u\nabla^2 h, \tag{23}$$

且初始条件为

$$t = T : u = u_T, \tag{24}$$

从而沿特征线的解为

$$u(\mathbf{x}, t) = u_T \exp\Big(-\int_t^T \nabla^2 h\big(\mathbf{x}(\tau; \mathbf{x}_T)\big)\mathrm{d}\tau\Big). \tag{25}$$

对 (21) 和 (25) 分别作离散化，得到迭代格式 (注意，这里沿时间 t 反向计算)

$$\mathbf{x}^{m+1} = \mathbf{x}^m + \nabla h(\mathbf{x}^m)\Delta t, \tag{26}$$

$$u^{m+1} = u^0 \exp\Big(-\sum_{j=1}^{m} \nabla^2 h(\mathbf{x}^j)\Delta t\Big), \tag{27}$$

其中 u^0 为 T 时刻污染元素的初始浓度.

若对某一给定点$\mathbf{x}^0$的浓度为u^0，则由(26)及(27)可以求出各时刻$t(t \le T)$沿着经过点$\mathbf{x}^0$的特征线上的浓度值$u(\mathbf{x}, t; \mathbf{x}^0)$．这样，当给定污染浓度的阈值$U$时，对给定时刻$\bar{t}(\bar{t} \le T)$浓度超过阈值$U$的区域为

$$B_{\bar{t}} = \{\mathbf{x}(\bar{t}) \mid u(\mathbf{x}(\bar{t}), \bar{t}; \mathbf{x}^0) \ge U\}, \tag{28}$$

此区域即为可能的污染源所在的区域．

6　评阅中发现的问题

在今年的评阅过程中，我们发现同学们大多能利用网络搜索资料，查阅了有关的参考文献，对数学软件(如Matlab等)的使用较为熟练．但同时也发现了一些问题．

1)　对数据的插值拟合，绝大多数参赛队都采用了软件来处理，但大多数队没有说明应该用什么方法，为什么用这种方法．多数队用了Matlab中的griddata函数来作插值，该函数提供了4种插值方法，而缺省方法是分块线性插值．分块线性插值得到的结果最主要的问题是不光滑，在后续处理时，无法求导，因此这是一种不太好的方法．

目前数学软件的功能非常强大，可以很方便地利用数学软件来处理我们的问题．但应注意，在使用时应该先了解所用的方法，特别注意该方法的适用范围，能否用于我们的问题，有什么局限性，可能会出现什么问题，然后再有的放矢地加以运用．

2)　对评价指标的选择，有些队只给出了单个指标，而未作综合处理．有些队仅对单个指标采用简单的比例处理．这样，对功能区评价及对整个区域的评价缺乏有力的数据支持，得到的结论比较粗糙．

绝大多数队都从网上查找现成的指标，但很少有同学深入研究这些指标．例如地质累积指数，这一指数未能充分利用问题给出的已知信息．题目中除了给出各重金属元素背景值的平均值，同时还给出了其标准差及范围．简单计算可知，其范围的上下限分别为$\mu_j \pm 2\sigma_j$，其中μ_j为第j种重金属元素背景值的平均值，σ_j为相应的标准差．因此在构造评价指标时，应该将标准差考虑进去．

一种改进的方法为对数据作标准化处理

$$\widetilde{c} = \frac{c - \mu}{\sigma}, \tag{29}$$

其中c为污染浓度的测量值，μ, σ分别为背景值的平均值与标准差，则改进的污染指数为

$$I = \log_2 \max(\widetilde{c} - 1,\ 1). \tag{30}$$

3)　相关性分析也可以用秩相关系数，方法是对两个指标分别按大小进行排序，记其序号分别为$1, 2, \cdots, n$，则秩相关系数定义为

$$\rho = 1 - \frac{6}{n(n^2-1)} \sum d_i^2, \tag{31}$$

其中d_i为指标之间的序号之差．

4)　污染元素的传播还可以用对流扩散模型与扩散模型建模．虽然我们已经说明了，在这个问题中扩散效应是很小的，但在确定污染源时，很多队假设污染浓度服从高斯分布，这实际上是用了单个点污染源的扩散模型，这一模型的解就是高斯分布．但对实测数据作正态性检验发现，污染浓度并不是正态的．因此，应该考虑多个点源的情况．

有的队在确定污染源时，对数据作了仔细的分析，发现污染浓度随着地形高度的增加是递减的，经回归分析发现，是指数衰减的．将高度影响剥除后，再用高斯正态分布的迭加形式进行搜索，这样得到的结果比较有说服力．

有的队在数据处理之前去除了浓度的异常值，这是不对的．异常值附近极有可能是污染源所在之处，将其去除，整体变得平缓了，污染源的位置也就变得不清晰了．

5) 当然，由于信息不足，题目只给出了一个时间点上的截面数据，没有时间方向上的序列数据，因而要精确确定污染源的位置是不可能的，需要作进一步的数据采集工作．但这是实际工作中经常碰到的情况，采集数据需要大量的人力、物力、财力，而且第一次采样前的数据是无法获得的．数学模型的重要作用就是在有限数据，甚至数据不足、有误差的情况下，给出合理的结果，以指导今后的工作．当然，数学模型也有其局限性，只能起到辅助的作用．

参考文献：

[1] 叶其孝. 大学生数学建模竞赛辅导教材(二)[M]. 湖南：湖南教育出版社，1997
Ye Q X. Teaching Materials for Undergraduate Mathematical Contest in Modeling (2)[M]. Hunan: Hunan Education Press, 1997

[2] 吴宗敏. 散乱数据拟合的模型、方法和理论[M]. 北京：科学出版社，2007
Wu Z M. Models, Methods and Theories of Scattered Data Fitting[M]. Beijing: Science Press, 2007

[3] 百度百科. http://baike.baidu.com/view/5817487.htm, 2011
Baidu Encyclopedia. http://baike.baidu.com/view/5817487.htm, 2011

[4] 柴世伟. 地累积指数法在土壤重金属污染评价中的应用[J]. 同济大学学报(自然科学版)，2006, 34(12): 1657-1661
Chai S W. Applications of cumulative index method in the evaluation of soil heavy metal pollution[J]. Journal of Tongji University (Natural Science), 2006, 34(12): 1657-1661

[5] 百度百科. http://baike.baidu.com/view/1693176.htm, 2011
Baidu Encyclopedia. http://baike.baidu.com/view/1693176.htm, 2011

[6] 百度文库. http://wenku.baidu.com/view/a2f5e04ae45c3b3567ec8b9f.html, 2011
Baidu Library. http://wenku.baidu.com/view/a2f5e04ae45c3b3567ec8b9f.html, 2011

[7] Anderson T W. An Introduction to Multivariate Statistical Analysis[M]. New York: Wiley-Interscience, 2003

[8] 张尧庭. 广义相关系数及其应用[J]. 应用数学学报，1978, 1(4): 312-320
Zhang Y T. Generalized correlation coefficient and its applications[J]. Acta Mathematicae Applagatae Sinica, 1978, 1(4): 312-320

[9] Ockendon L, *et al.* 应用偏微分方程[M]. 北京：科学出版社，2008
Ockendon L, *et al.* Applied Partial Differential Equation[M]. Beijing: Science Press, 2008

[10] 蔡志杰，龙敏斌. 城市表层土壤中重金属污染的传播模型[J]. 数学建模及其应用，2012, 1(1): 53-56
Cai Z J, Long M B. Propagation model of heavy metals contamination in urban topsoil[J]. Mathematical Modeling and its Applications, 2012, 1(1): 53-56

Analysis of Heavy Metals Contamination in Urban Topsoil

CAI Zhi-jie[1,2]

(1- School of Mathematical Sciences, Fudan University, Shanghai 200433;
2- Shanghai Key Laboratory of Modern Applied Mathematics, Shanghai 200433)

Abstract: In this paper, we discuss Problem A in CUMCM 2011 and present the methods of solving these problems. We also give some comments on the students' papers.

Keywords: heavy metal contamination; interpolation for scattering data; evaluation index; convection equation; character method

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0050-09

城区重金属污染评价及源识别模型研究

齐 谦， 成生辉， 燕斌斌
指导教师： 刘保东
(山东大学，济南 250100)

编者按: 本问题的数据来源于某城市对土壤环境的实地检测，现实条件的限制为建立好的数学模型提出挑战．因此，好的数学模型就需要结合实际情况，充分利用已经获得的数据，在合理假设下尝试寻找尽可能好的解决方案．本论文对于背景值的处理是一个亮点，由于题目给出的背景值除均值外，还给出了标准差，如何利用这个信息去获得更为可靠的评价标准是需要考虑的，而不是机械的套用书本．对于第三问，由于数据只给出了散乱位置在一个时间点的污染物浓度，因此就需要利用这些有限的信息建立数学模型去获得尽可能可靠的污染源位置估计，本论文在合理假设的基础上，给出了两种可行的估计模型，以及在这些模型基础上的综合分析方法，因此可以为判断污染源提供好的参考．

摘 要: 本文首先根据问题给定的散乱数据特点，构建了基于Delauney三角网格剖分的三角样条插值模型，得到了各污染物的空间分布，同时针对背景值浓度是一个区间的特点，分别建立了相应的单因子评价模型和内梅罗污染指数综合评价模型，据此确定了各不同区域的污染程度．同时分别利用多元聚类和相关性分析，给出了各重金属元素之间的关联程度，据此确定了各重金属污染的主要原因．对污染源位置的识别，分别构建了基于三角形样条插值、对流扩散和分布函数的模型，最后给出了一种组合模型优化求解的方法，通过模型之间的互补，给出源位置识别的结果．

关键词: 三角样条插值；聚类分析；相关性分析；对流–弥散方程；正态分布拟合；均匀分布拟合

分类号: AMS(2000) 42A15; 35K57 **中图分类号:** X28; O29 **文献标识码:** A

1 问题提出

人类的生产和生活活动是造成环境污染的主要因素[1]．为了调查某城市城区表层土壤中重金属污染的现状，首先将所考察的城区划分为间距为1公里左右的网格子区域，然后按照每平方公里1个采样点的方式对表层土(0–10厘米深度)进行取样、编号，并用GPS记录采样点的位置．对所取样品应用专门仪器测试分析，从而获得了319个样本所含的多种化学元素的浓度数据．另一方面，按照2公里的间距在那些远离人群及工业活动的自然区取样，将其作为该城区表层土壤中元素的背景值．为便于区分，将城区按照功能划分为生活区、工业区、山区、主干道路区及公园绿地区等，分别记为1类区、2类区、···、5类区，不同的区域环境受人类活动影响的程度不同．现需要根据题目所给数据解决下面四个问题：

1) 给出8种主要重金属元素的空间分布，并分析该城区内不同区域重金属的污染程度；

2) 通过数据分析，说明重金属污染的主要原因；

3) 分析重金属污染物的传播特征，由此建立模型，确定污染源的位置；

4) 分析所建立模型的优缺点，为更好地研究城市地质环境的演变模式所应收集何种信息提供依据，另外，基于这些信息针对性的建立模型解决问题．

2 基本假设

1) 各个采样点属于单一功能区，而且数据准确；

2) 各区域土壤中重金属的背景值相同，不考虑区域功能差异和地形差异；
3) 现有重金属的监测结果是污染物常年沉积形成的；
4) 污染源的位置在局部浓度最高点附近；
5) 当监测浓度值大于背景值的上限时，即认为被污染.

3 问题一的模型建立与求解

3.1 8种主要重金属元素在该城区的空间分布

利用319个区域监测点的观测结果，确定其空间分布，即要估计区域内任一点(x,y)处的浓度值. 由插值理论可知，该问题本质上是一个二维插值问题. 但由于实地人工测量时存在局限性，本题给出的样本观测点在空间中呈现散落且无规律的特征，显然利用一般的二维矩形网格插值方法并不可行. 为此，通过文献检索，本文采用一种基于Delauny三角剖分的三角样条多项式插值方法，对已知的散乱数据进行三角样条多项式插值. 该方法的数学描述如下[2,3].

设T是三角形网格中具有顶点$v_l=(x_l,y_l)$, $l=i,j,k$的三角形，F_l是顶点v_l, $l=i,j,k$处的实际观测值(浓度或海拔高度). 构造基于该三角形的样条插值曲面函数$F_T(x,y)$:

$$F_T(x,y)=W_i(x,y)F_i+W_j(x,y)F_j+W_k(x,y)F_k, \tag{1}$$

其中$W_i(x,y)$, $W_j(x,y)$, $W_k(x,y)$是权函数，定义为

$$W_i(x,y)=\Big[10L_i-15L_i^2+6L_i^3+30L_jL_k\Big(\frac{Q_kL_k}{D_j}+\frac{Q_jL_j}{D_k}\Big)\Big]L_i^2,$$

$$W_j(x,y)=\Big[10L_j-15L_j^2+6L_j^3+30L_iL_k\Big(\frac{Q_kL_k}{D_i}+\frac{Q_iL_i}{D_k}\Big)\Big]L_j^2,$$

$$W_k(x,y)=\Big[10L_k-15L_k^2+6L_k^3+30L_kL_j\Big(\frac{Q_iL_i}{D_j}+\frac{Q_jL_j}{D_i}\Big)\Big]L_k^2,$$

式中S为三角形面积

$$L_i(x,y)=\frac{1}{2S}[x_jy_k-x_ky_j+(y_j-y_k)x+(x_k-x_j)y],\quad D_i=(x_k-x_j)^2+(y_k-y_j)^2,$$

$$L_j(x,y)=\frac{1}{2S}[x_ky_i-x_iy_k+(y_k-y_i)x+(x_i-x_k)y],\quad D_j=(x_i-x_k)^2+(y_i-y_k)^2,$$

$$L_k(x,y)=\frac{1}{2S}[x_iy_j-x_jy_i+(y_i-y_j)x+(x_j-x_i)y],\quad D_k=(x_j-x_i)^2+(y_j-y_i)^2,$$

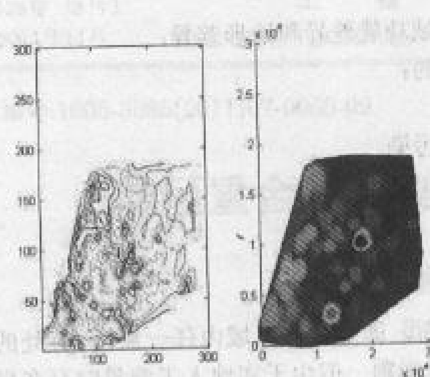
$$Q_i=(x_j-x_i)(x_k-x_i)+(y_j-y_i)(y_k-y_i),\quad Q_j=(x_k-x_j)(x_i-x_j)+(y_k-y_j)(y_i-y_j),$$

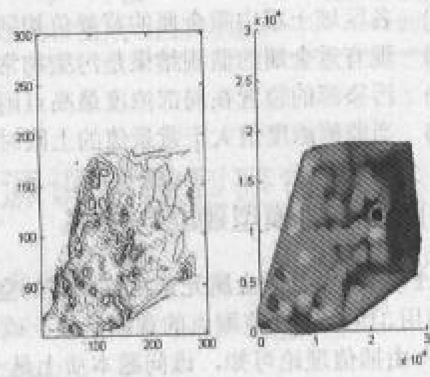
$$Q_k=(x_i-x_k)(x_j-x_k)+(y_i-y_k)(y_j-y_k).$$

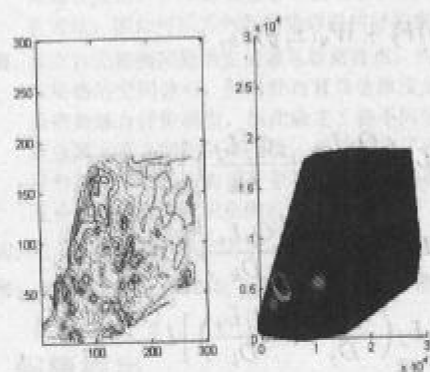
可以证明$F_T(x,y)$具有如下性质[2,3]:

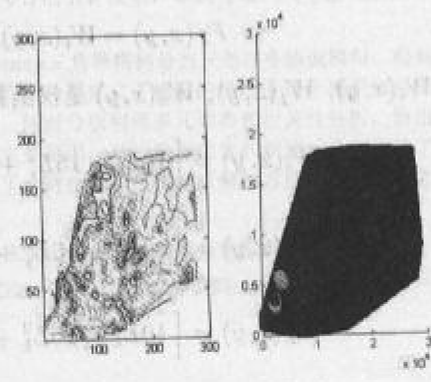
a) 在T的顶点$v_l=(x_l,y_l)$处，$W_l(x,y)=1$, $l=i,j,k$;
b) $W_l(x,y)$在v_l的对边上的函数值为零，在T的边界上，其一阶法向偏导数为零；
c) $W_i(x,y)+W_j(x,y)+W_k(x,y)=1$.

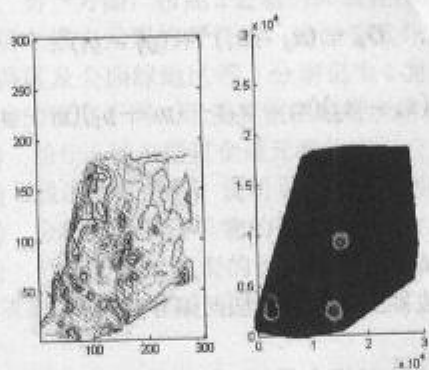
通过给定的采样点的数据，利用上述三角样条插值法对样本数据进行插值，得出拟合后的8种重金属的空间分布，见图1至图8所示(左侧为城区等高线图，右侧为污染物浓度空间分布图，颜色从蓝到红，由浅到深表示污染物浓度升高).

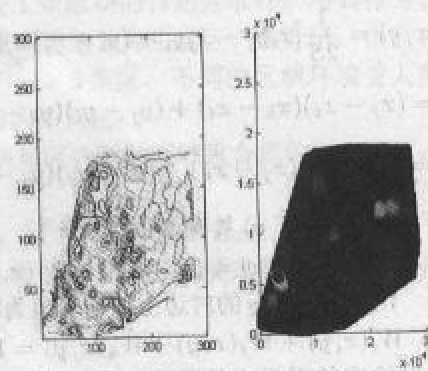

图 1: As空间分布


图 2: Cd空间分布


图 3: Cr空间分布


图 4: Cu空间分布


图 5: Hg空间分布


图 6: Ni空间分布

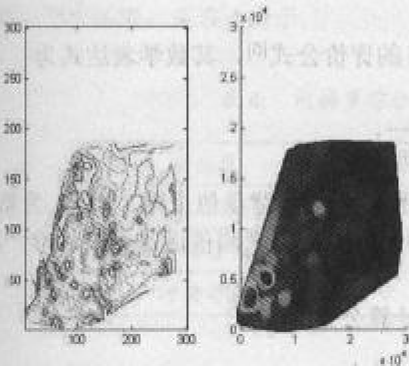


图 7:　Pb 空间分布

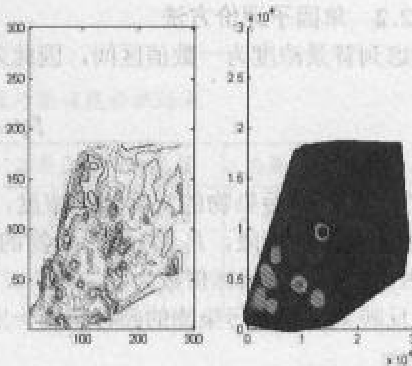


图 8:　Zn 空间分布

综合分析拟合浓度数据和污染物浓度空间分布图与城市区域分布图，分析结果，如表 1.

表 1:　8 种重金属空间分布分析表

As	含量普遍较高，最大值出现在交通区，说明主干道路污染整体不大，但个别区域污染较严重
Cd	除山区外，该城区大部分区域Cd含量超标，在工业区、交通区，超标现象明显
Cr	总体分布较集中，高浓度区主要出现在工业区，生活区和交通区也有较集中的分布
Cu	总体分布很集中，主要出现在工业区和交通区的交界处
Hg	分布很集中，主要出现在交通区和工业区，最高值出现在交通区
Ni	分布很集中，集中分布在主干道路和工业区，最大值出现在交通区
Pb	集中分布在工业区和交通区交界的两点，生活区和公园绿地区的含量也超标
Zn	集中分布在七个点上，在工业区、交通区和生活区有较明显的分布，最大值出现在工业区

3.2　不同区域重金属污染程度评价

3.2.1　评价标准的确定

鉴于题目给出的各重金属的背景浓度监测结果是一个统计结果，考虑到背景值可直观地认定为自然状态下的土壤重金属自然含量，只要观测结果满足背景值范围，即可认为未受污染，因此可直接作为评价标准[4,5]，如表 2 所示.

表 2:　评价标准 (单位: ug/g)

元　素	As	Cd	Cr	Cu
标准限值	1.8–5.4	0.07–0.19	13–49	6.0–20.4
元　素	Hg	Ni	Pb	Zn
标准限值	0.019–0.051	4.7–19.9	19–43	41–97

3.2.2 单因子评价方法

考虑到背景浓度为一数值区间，因此采用类似于pH的评价公式[6]，其数学表达式为

$$P_i = \frac{C_i - C_{\min_i}}{C_{\max_i} - C_{\min_i}}, \tag{2}$$

其中 C_i 为第 i 种污染物的实际监测浓度，$C_{\max_i}$ 为第 i 种污染物的背景值上限，$C_{\min_i}$ 为第 i 种污染物的背景值下限，P_i 第 i 种污染物的污染指数．如 $P_i \le 1$，则观测值满足评价标准，不超标；否则，超标，超标倍数为 $P_i - 1$．

为反映某区域某污染物的超标范围，定义超标率．计算公式如下

$$q = \frac{m}{n} \times 100\%, \tag{3}$$

其中 m 为该区域中监测结果超标的点位个数，n 为区域中所设的监测点的总数．

3.2.3 单因子评价结果统计与分析

通过对不同区域各采样点的监测结果的单因子评价与结果统计，得到不同区域各重金属元素的统计指标，包括：采样点数，超标点数，超标率，平均值，最大超标倍数等．从统计结果可以看出：该城区各不同功能区域中8种重金属元素土壤监测结果均有超标现象，说明该区域已不同程度地受到人类活动的影响．其中生活区超标率最高的是Cr、Cu、Zn、Hg，最大超标污染物为Zn和Hg，分别超出背景值标准49.94和19.31倍．工业区各重金属元素超标都很严重，尤其是Cu、Cd、Pb、Hg、Cu，最大超标污染物为Hg和Cu，分别超出背景值标准420.28和174.17倍．山区所有重金属超标点率均在20%左右，土壤环境质量相对较好．主干道路区除Cd、Ni外，其余超标均很严重，主要超标污染物为Cu、Zn、Hg，最大超标污染物为Hg和Cu，分别超出背景值标准的498.41和93.36倍．公园绿地区超标情况也不乐观，主要超标污染物为As、Cu、Hg、Zn，最大超标污染物为Hg和Zn，分别超出背景值标准40.26和23.08倍．

3.2.4 不同区域污染程度综合评价

1) 评价方法：对每一个功能区，采用内梅罗综合污染指数法．计算公式如下[2]

$$P_z = \sqrt{\frac{P_{av}^2 + P_{\max}^2}{2}}, \tag{4}$$

其中 P_z 为综合污染指数(综合反映各污染物对区域土壤的不同作用)，P_{av} 为所有单项污染指数的平均值，$P_{\max}$ 为土壤环境中各单项污染指数中的最大值．

2) 污染指数分级方法：利用参考文献[4]，确定分级方法，如表3所示．

表3: 综合污染指数的分级标准

污染等级	P_z	污染程度	污染描述
I	$P_z \le 0.7$	安全	清洁
II	$0.7 < P_z \le 1.0$	警戒线	尚清洁
III	$1.0 < P_z \le 2.0$	轻污染	超标
IV	$2.0 < P_z \le 3.0$	中污染	土壤、作物受中度污染
V	$P_z > 3.0$	重污染	土壤、作物受严重污染

3) 评价结果：见表4所示.

表4: 内梅罗综合污染指数评价及污染程度分级结果

功能区	1类区	2类区	3类区	4类区	5类区
平均指数值	2.05	5.19	0.69	3.54	1.62
内梅罗综合指数值	3.5	19.48	0.79	13.37	2.02
污染等级	V	V	II	V	IV

4) 结果分析：从表4中可以看出，评价城区中，污染最重的区域为2类区(工业区)，其次是4类区(主干道路区)，1类区(生活区)和5类区(公园绿地区)属中度污染，3类区(山区)污染最轻，属尚清洁区域．总体来说，该城区的重金属污染较为严重．

4 问题二的模型建立与求解

首先通过观察图1至图8及表3的数据评价结果，发现主要超标污染物和最大超标污染物呈一定规律性，这说明不同重金属之间很可能存在很高相关性，即可能来自相同的污染源．为了定量分析各重金属元素之间的关联性，本文利用多元聚类分析对不同重金属进行相关性分析，考察污染物之间的相关性．另外利用一种方法给出不同因子之间的相关系数．若重金属含量有显著的相关性，说明可能来源相同，这与污染物产生的原因是密切相关的，由此对照不同污染源的污染特征，可大致确定出产生这些污染的原因．最后通过观察地形地貌图，结合海拔高度分布值，给出不同重金属污染的原因．

聚类分析法所采用的相关性度量标准为：

$$d(\mathbf{x},\mathbf{y})=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}, \tag{5}$$

其中$d(\mathbf{x},\mathbf{y})$表示向量$\mathbf{x},\mathbf{y}$的距离．

相关性系数法所采用的相关性度量标准为：

$$r=\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum(x_i-\overline{x})^2(y_i-\overline{y})^2}}. \tag{6}$$

通过计算，得各功能区主要元素相关性分析结果，如表5所示．

表5: 各区显著相关或高度相关的重金属

功能区	1类区	2类区	3类区	4类区	5类区
主要相关元素	Cd–Pb	Cr–Cu Cr–Hg Cu–Hg	Cr–Ni	Cr–Cu Cr–Ni Cu–Ni	Cr–Ni Cu–Pb Pb–Zn

按区域分析污染的主要原因：

1) 各个功能区中，超标率最高的为Cd和Cu，可见该区域是Cd和Zn的高污染区. 形成污染的原因有很大可能性来源于生活垃圾填埋或任意排放及含铜农药喷洒所致，也可能来源于相关冶炼企业的烟尘和垃圾排放.

2) 生活区中，Cd与Pb高度相关，说明它们污染源相似，可能与生活垃圾中的废旧电池、电路板及汽车尾气排放有关.

3) 工业区中，几乎所有的重金属都显著相关，说明它们污染源相似，而且污染程度明显高于其他功能区. 可能与工业烟尘和工业垃圾填埋有关. 此外Hg的含量明显超标，可能存在含Hg工业垃圾填埋或燃煤电厂粉尘排放.

4) 山区中，各污染物污染指数均较低，Cr与Ni显著相关，说明该区污染主要来源于生活或工业垃圾污染.

5) 交通区中，除Hg之外，其余7种重金属具有显著相关性，说明，交通区的Hg来源比较分散，由问题一可知，该区主要超标污染物为Cd 、Hg、Zn，其中Ni、Pb、Cd应主要来自汽车尾气、轮胎磨损以及大气降尘.

6) 公园绿地区中，虽然超标指数较低(高于山区)，但是几乎各个指数都超过标准. 造成污染的原因可能是因为废气或者用废水灌溉或者农业(农药、化肥)造成的.

5 问题三的模型建立与求解

模型一 基于空间散乱数据的三角样条多项式插值的污染源识别模型

首先根据题目给定的散乱观测点，构造基于Delauny三角剖分的三角形剖分网格. 其次，在三角形网格上分别构造基于浓度和海拔高度的二元三次插值多项式，如式(1)所示.

通过三角样条函数能够分别给出地形高度函数和浓度空间分布函数，通过求导或局部加密网格计算，可以直接得到浓度最高点的坐标值及海拔高度，即污染源的坐标[7]. 识别结果见表6.

表6: 污染源位置识别结果

重金属名称	污染源坐标
As	A(18210,10060),B(12650,3023),C(27500,12090),D(4600,7404)
Cd	A(21470,11410),B(4504,11350),C(2300,3085)
Cr	A(3285,5862)
Cu	A(2300,3517)
Hg	A(2683,2406),B(15240,9194),C(13800,2345)
Ni	A(3354,5862)
Pb	A(2108,3332),B(2792,4998)
Zn	A(13800,9626),B(9487,4443)

模型二 对流–扩散方程模型

雨水的作用是重金属在土壤表层中传播的主要原因之一，重金属在土壤中主要以溶质的形式存在，而其传播方式主要以对流扩散方式为主. 由此建立对流扩散方程[8]：

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} - u\frac{\partial C}{\partial x}, \tag{7}$$

其中 $\frac{\partial C}{\partial t}$ 是浓度关于时间的导数，$D\frac{\partial^2 C}{\partial x^2}$ 为扩散项，它表明自然扩散的速率正比于浓度的梯度，D 为弥散系数，$u\frac{\partial C}{\partial x}$ 为对流作用产生的污染物浓度的迁移速率，u 为对流速度．在本文所研究的问题中，对流速度 u 可视为与地形高度梯度成正比的函数，即 $u = k\nabla h(x, y)$，$h(x, y)$ 为地形高度曲面函数(可由模型一近似得到)．

*在给定初始和边界条件下，式(7)的求解可由解析法或数值计算方法给出模型解．但在对模*型实际处理时，因无法确定初始、边界条件，因此直接求解模型存在实际上的困难．

模型三　分布模型

由文献检索可知[8]，标准的对流扩散方程存在解析解表达式．它的解一方面类似热传导方程，随着时间变化，浓度趋于均一；另一方面，与波动方程类似，一直向一个方向传播．本质上是一个正态分布模型．

正态分布模型(对流为主)

受地形的影响，污染物在不同方向上的分布是不同的．一般情况下，沿相对平缓的方向的分布是比较理想的正态分布，而如果两边出现障碍，可能会使分布更加向中心集中．因此，用各向相异的正态分布模型拟合，可以达到更准确的效果．

$$C(x, y) = C_0 e^{\lambda_1 (a_1(x-x_0)+b_1(y-y_0))^2 + \lambda_2 (a_2(x-x_0)+b_2(y-y_0))^2}, \tag{8}$$

其中 λ_1，λ_2，a_1，a_2，b_1，b_2 为待定参数，x_0，y_0 为污染源位置坐标，C_0 为污染源源强，$C(x, y)$ 为污染物在 (x, y) 处的浓度值．

均匀分布模型(扩散为主)

若污染物空间地形各方向的传播都受到限制，则污染物会在污染源附近堆积，这时可以将污染物的分布特征用均匀分布近似地表示出来

$$F(x, y) = ax^2 + bxy + cy^2 + dx + ey + f. \tag{9}$$

若 $F(x, y) < 0$，$(x, y) = C_0$；$F(x, y) > 0$，$C(x, y) = C_1$，其中 C_0 为污染源区域内部的点的浓度，C_1 为污染源外部的点的浓度．

模型求解方法：

1)　选定疑似污染源附近临近网格作为局部模拟区域，得到局部插值函数；

2)　加密局部网格，由插值函数计算各网格点的函数值，作为观测值；

3)　由非线性最小二乘方法，建立解析解与观测值的误差平方和，求解待定参数，使误差平方和最小．

$$\min f(\mathbf{a}, x_0, y_0) = \sum \left(C(x_i, y_i) - C(x_i, y_i, z_i)\right)^2, \tag{10}$$

其中 $\mathbf{a}$ 为待识别参数向量；

4)　识别污染源位置：所求的 (x_0, y_0)，即为该污染源的平面位置．再由地形曲面插值函数，计算其空间海拔高度，即可确定污染源的实际空间位置．

通过模拟、仿真发现，求解结果与模型一基本相近．

6　问题四的模型建立与求解(略)

参考文献：

[1] 管东生，陈玉娟，阮国标. 广州城市及近郊土壤重金属含量特征及人类活动的影响[J]. 中山大学学报，2001, 40(4): 93-96

Guan D S, Chen Y J, Ruan G B. Study on heavy metal concentrations and the impact of human activity on them in urban and suburb soils of Guangzhou[J]. ACTA Scientiarum Naturalium Universitis Sunyatseni, 2001, 40(4): 93-96

[2] 张彩明，孙德法，汪嘉业. 散乱数据点的三次多项式插值[J]. 计算机辅助设计与图形学学报，1998, 10(5): 416-424

Zhang C M, Sun D F, Wang J Y. Cubic polynomial interpolation of scattered data points[J]. Journal of Computer-Aided Design & Computer Graphics, 1998, 10(5): 416-424

[3] 聂志峰，等. 基于 C^1 自然临近插值的曲面拟合[J]. 工程图学学报，2010, 1: 110-115

Nie Z F, *et al.* Surface fitting based on C^1 natural neighbor interpolation[J]. Journal of Engineering Graphics, 2010, 1: 110-115

[4] 姜芝萍，杨俊衡. 城市重点污染区土壤重金属污染评价标准探讨—以衡阳市某区为例[J]. 安全与环境工程，2010, 17(1): 57-64

Jiang Z P, Yang J H. Discussion on heavy metal quality assessment standards of urban pollution soil—taking Hengyang city for example[J]. Safety and Environmental Engineering, 2010, 17(1): 57-64

[5] 夏立江，王宏康. 土壤污染及其防治[M]. 上海：华东理工大学出版社，2001

Xia L J, Wang H K. Soil Pollution and Prevention[M]. Shanghai: East China University of Science and Technology Press, 2001

[6] 余剑东，倪吾钟，杨肖娥. 土壤重金属污染评价指标的研究进展[J]. 广东微量元素科学，2002, 9(5): 11-17

Yu J D, Ni W Z, Yang X E. Current progress in evaluation indexes for heavy metals pollution in soils[J]. Guangdong Trace Elements Science, 2002, 9(5): 11-17

[7] 周博，等. Matlab 科学计算[M]. 北京：机械工业出版社，2010

Zhou B, *et al.* Scientific Computing with Matlab[M]. Beijing: China Machine Press, 2010

[8] 宋新山，吴应玲，陈栋. 基于 Matlab 的对流弥散方程的不稳定源解[J]. 环境科学与技术，2008, 31(11): 31-34

Song X S, Wu Y L, Chen D. Numerical simulation for volatile source terms in convection-dispersion equation based on Matlab[J]. Environmental Science & Technology, 2008, 31(11): 31-34

Study on Urban Heavy Metal Pollution Assessment and Source Recognition Model Assessment

QI Qian,　CHENG Sheng-hui,　YAN Bin-bin

Advisor:　LIU Bao-dong

(Shandong University, Jinan 250100)

Abstract: According to the scattered features of the given data, a triangle spline interpolation method is constructed based on the Delauney triangular subdivision technique. By utilizing this method, the spatial distribution for each heavy metal is then obtained. Based on the fact that the background value corresponds to an interval for each heavy metal, the factor assessment model and the comprehensive evaluation method are established. The assessment result for each subregion is also provided. Besides, by using the multivariate clustering and correlation methods, the relationships between different heavy metals is obtained, and the main reason underlying the urban heavy metal pollution is evaluated. To recognize the source location of the heavy mental pollution, three models are constructed respectively. The first one is established based on the triangular spline interpolation technique, the second based on the characteristic of convection-dispersion, and the third based on the distribution characteristics of convection or dispersion. By combining these three models, the method for recognizing the source location is proposed and its effectiveness is substantiated by experiments.

Keywords: triangle spline interpolation; clustering analysis; correlation analysis; convection–dispersion equation; normal distribution fitting; evenly distributed fitting

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0059-08

基于空间数据的城市土壤重金属污染分析

盛振峰， 黄 思， 周敏强
指导教师： 曹春正
（南京信息工程大学，南京 210044）

编者按：本文源于2011年高教社杯全国大学生数学建模竞赛本科组高教社杯获得者的参赛论文．

摘 要：随着工业迅速发展，重金属污染物对环境产生了巨大的影响．本文利用地累积污染指数衡量重金属污染程度，同时利用污染负荷指数比较不同功能区的污染程度，按污染指数分布特征将重金属归为四类，结合重金属来源的国家分类标准分析了各种元素可能的污染来源．对空间数据进行指数衰减高程订正，利用加权混合二元正态分布密度函数拟合多污染源传播形成的浓度曲面，确定了重金属污染源的具体位置．

关键词：重金属；污染分析；地统计分析；空间数据；地累积指数；混合正态分布

分类号：AMS(2000) 86A32 中图分类号：X53 文献标识码：A

1 问题的重述(略)

2 模型假设

1) 8种元素的扩散方式相同，并且不互相干扰；
2) 只考虑土壤表层污染，不考虑重金属污染在地面以下向地心方向的扩散．

3 符号说明

i:	重金属种类；	S_i:	第 i 种金属实测值浓度的标准差；
I_{geo}:	地累积指数；	K:	衰减因子；
C_i:	第 i 种重金属的实测浓度；	$J(h)$:	浓度随高度扩散的函数；
B_i:	第 i 种重金属的环境背景浓度；	$I(x_i,y_i)$:	浓度随平面位置变化的函数；
A:	背景值变动系数．		

4 对问题一的分析和处理

4.1 对每种金属的污染情况分析

本文采用Muller地累积指数刻画重金属污染程度[1]，对城区重金属污染分布和污染程度进行分析．地累积指数考虑了人为污染因素、环境背景值、自然引起的背景值变动，并且还能对污染程度进行分级．地累积指数的计算式如下：

$$I_{\text{geo},i}=\log_2\left[C_i/(A\times B_i)\right], \tag{1}$$

其中 A 为考虑各地沉积岩差异可能引起的背景值变动系数．根据 $I_{\text{geo},i}$ 的值，规定了相应的污染程度级别划分标准，见表1．

表 1:　Muller 地累积指数污染物等级划分

I_{geo} 指数	< 0	0–1	1–2	2–3	3–4	4–5	> 5
级数	0	1	2	3	4	5	6
累计污染程度	无	无–中	中	中–强	强	强–极强	极强

由于处理的数据区域并不是规则的矩形，本文利用 ArcGIS 9.3，把 1km 分辨率的观测值利用克里金插值插入到 73m 分辨率的空间网格点中，绘制出城区海拔高度、功能区以及各重金属污染等级分布图与浓度指数分布图，以 Hg 为例，如图 1 和图 2.



图 1:　Hg 的污染分级图

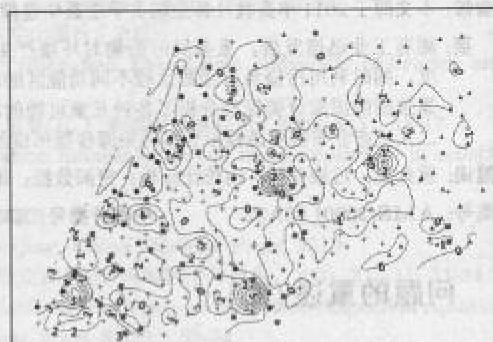


图 2:　Hg 的污染指数图

综合考虑城区地形以及各个元素的污染分级图和污染指数图，对 8 种元素土壤浓度的空间分布做讨论后，得到表 2.

表 2:　污染指数分布的大致趋势

重金金属	地累积指数	污染分布
Ni	(−0.75, 0.15)	东部和中部无污染，西南地区轻度–中级污染
Cr	(−0.8, 1.6)	东部和中部无污染，西南地区轻度–中级污染到中度污染
As	(−0.80.6)	东部和中部无污染，西部轻度–中级污染，但污染区域均匀且范围较大
Cd	(−0.8, 1.2)	东部无污染，中部和西部轻度–中级污染，西南角中度污染
Pb	(−0.6, 1.6)	东部无污染，中部和西部轻度–中级污染，西南角中度污染
Zn	(−0.82.2)	东部无污染，中部轻度–中级污染，西部中度污染，西南角中强度污染
Cu	(−0.12, 2.8)	东北角无污染，东部和中部轻度–中级污染，西部中度污染，西南角中强度污染
Hg	(−0.2, 5)	东部无污染，中部有散布的轻度–中级污染，西部中度污染，呈多峰值散乱分布

4.2　对每个功能区的污染程度分析

地积累指数分析了每种重金属的污染分布，但不同污染物、污染源有不同的环境效应、不同的危害，为了评定不同功能区的重金属的综合污染程度，使之能在同一尺度上比较，引入污

染负荷指数的分析各功能区的污染程度

$$\widetilde{C}=\frac{C_i}{C_{0i}},\quad PLI=\sqrt[n]{\prod_{i=1}^{n}\widetilde{C}_i},\quad PLI_{zone}=\sqrt[m]{\prod_{j=1}^{m}PLI_j},\tag{2}$$

其中C_i为第i种金属浓度的实测值，C_{i0}第i种金属的背景值；PLI为单个点的污染负荷指数，n为金属污染物种数；PLI_{zone}为区域的污染负荷指数，m为采样点个数．根据PLI的值，规定了相应的污染负荷指数等级划分标准，见表3.

表3: 污染负荷指数等级划分

PLI值	<1	1–2	2–3	>3
污染等级	0	1	2	3
污染程度	无污染	中等污染	强污染	极强污染

利用式(2)计算得到工业区、主干道区、生活区、公园绿地区、山区的PLI_{zone}分别为2.35, 1.92, 1.83, 1.58, 1.06，即五大功能区的污染程度为：工业区 > 主干道区 > 生活区 > 公园绿地区 > 山区．这一结果符合生活实际，工业区污染最强，其余四个功能区的污染程度都为中等污染，但是山区更接近无污染，主干道更接近于强污染．

5 对问题二的分析和处理

本文的假设所有的重金属扩散模式相同，因此若重金属离子具有类似的分布，可以认为它们具有相同的来源．根据问题一中对每种元素污染指数分布的分析，将分布相同或类似的元素合并成4个大类，如表4.

表4: 每一大的类型重金属污染分布特征

类型	元素	特征
1	Ni, Cr	污染程度最低，东部和中部没有污染， 西部少量地区有轻微污染，污染范围小
2	As	污染程度较低，东部和中部无污染， 西部轻度–中级污染，但污染区域均匀且范围较大
3	Cd, Pb, Zn, Cu	污染程度中等偏高，污染等级呈阶梯分布， 受中等污染的范围较大，个别地区出现强污染
4	Hg	污染程度较高，污染分布散乱，似奶牛斑， 总体污染程度较低，但在某些点出现极强污染

下面运用统计方法[2-4]分别针对标准化浓度和地累积指数验证表4分类的合理性．标准化数据浓度是对原始浓度标准化后得到的，它能反映重金属浓度空间分布的实际情况．地累积指数则是考虑到了重金属浓度的背景值信息，侧重于评定重金属污染等级．因此这两种数据类型各有侧重．

利用SPSS 18.0，计算8种元素的地累积指数的简单相关系数矩阵，得到Zn, Pb, Cd, Cu之间存在非常大的相关性，Cr和Ni的相关性比较大，As和Hg是两种比较独立的元素，与其他几种因素指数的空间分布相关性不是非常显著.

对标准化浓度数据的相关分析可得，Cr和Ni之间有显著的相关性，Pb和Cd、Pb和Cu，以及Pb和Zn之间相关性也比较显著，而As和Hg与其他元素的相关性相对较小，属于两种比较独立的元素. 地累积指数和标准化浓度数据相关分析的结论相互印证，说明了地累积指数在该问题中能够较好的反映实际的污染情况，另外也证明了表4分类的合理性.

利用SPSS 18.0，对8种重金属空间分布采用主成分分析法分析污染成因，结果表明：6种重金属Cd, Cr, Cu, Ni, Pb, Zn可能有一个相同的来源. As和Hg比较独立.

综上可以得到：Ni和Cr相关非常显著，Pb, Zn, Cd, Cu的相关关系显著，我们假定它们分为独立的两类，并且每一类具有相同的污染来源，而As和Hg为比较独立的两种元素，因此认为它们源于不同的污染源. 从《中国土壤重金属来源》可以总结出Pb主要源自工业、尾气；Cu主要源自有工业、农药；Ni, Cr, Cd, Zn主要源自于工业；As主要源自于矿业开采和农药；Hg的主要源自于废水和生活垃圾. 为了全面的分析题目所给区域中的污染形势，提炼出了污染分布5个关键区，即污染高值区，如图3，关键区中包含2个浓度异常区 (A区和E区). 对这些区域进行详细的分析，进一步给出了污染最可能的来源.

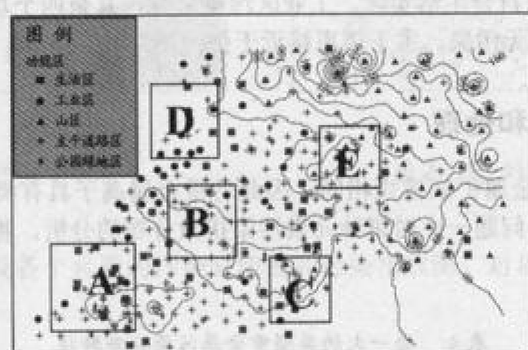


图3:　污染集中分布区域图

例如在B区，Pb, Cr, Cu, Zn都有高污染值. 根据国家根据《中国土壤重金属来源》，工业废渣、废水、废气都是这4种金属的来源，除此外Cu可以来源于农业，Pb还可以来源于汽车尾气. 但从图中可以看出B区主要是道路交通主干道，因此我们猜测该区Pb的一个来源可能是汽车尾气；而B左上角处为大片的工业区，但这4种重金属浓度反而不大，我们进一步猜测在B区有一集中的垃圾场，在此汇集的工业废物是4种金属的共同来源. 对其余几个关键区类似分析在此不再赘述.

6　对问题三的分析和处理

6.1　扩散模型

设有一污染源，沿x, y, z三个方向的扩散，扩散系数分别为a^2, b^2, c^2，设$u(x,y,z,t)$是t时刻点(x,y,z)处某物质的浓度. 任取一个闭曲面S，围成的区域为Ω，由扩散理论中的费克第

一定律，Δt 时间内，通过 S 流入 Ω 的质量为

$$M_1 = \int_t^{t+\Delta t} \iint_S \left(a^2 \frac{\partial u}{\partial x} \cos\alpha + b^2 \frac{\partial u}{\partial y} \cos\beta + c^2 \frac{\partial u}{\partial z} \cos\gamma \right) \mathrm{d}S\mathrm{d}t, \tag{3}$$

由于衰减，Δt 时间内 Ω 中的质量减少为

$$M_2 = \int_t^{t+\Delta t} \iiint_\omega k^2 u \mathrm{d}V\mathrm{d}t, \tag{4}$$

其中 k^2 为衰减系数[5]．Δt 时间内 Ω 内浓度变化引起的质量增加为

$$M_3 = \iiint_\Omega [u(t+\Delta t) - u(t)]\mathrm{d}V = \int_t^{t+\Delta t} \iiint_\Omega \frac{\partial u}{\partial t} \mathrm{d}V\mathrm{d}t. \tag{5}$$

由物质质量守恒定律，显然 $M_3 = M_1 - M_2$，于是得衰减的扩散过程的数学模型为

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + b^2 \frac{\partial^2 u}{\partial y^2} + c^2 \frac{\partial^2 u}{\partial z^2} - k^2 u. \tag{6}$$

设扩散源在点 (x_0, y_0, z_0) 处，则此扩散问题满足 Cauchy 问题．用傅立叶变换可求得解析解为

$$u(x,y,z,t) = \frac{M}{8\pi tabc\sqrt{\pi t}} \exp\left\{ -\frac{(x-x_0)^2}{4a^2 t} - \frac{(y-y_0)^2}{4b^2 t} - \frac{(z-z_0)^2}{4c^2 t} - k^2 t \right\}, \tag{7}$$

其中 M 为扩散源的质量．

由于在无限大介质中扩散浓度集中在点源，这类扩散问题都具有高斯分布的形式，即“高斯解”．由此受到启发，假设污染扩散浓度模式为多元的正态分布曲面．多元正态分布是一元正态分布的直接推广．设随机向量 $X = (x_1, x_2, \cdots, x_p)^\top$ 服从 p 维正态分布，则其密度函数为

$$f(X) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left\{ -(X-\mu)^\top \Sigma^{-1} (X-\mu) \right\}, \tag{8}$$

或可简化地表示为

$$g(u) = k \exp\left(-\frac{u}{2} \right), \tag{9}$$

其中 k 为污染源的权重系数，u 为马氏距离．

题目所给的浓度信息是包含高度信息的空间数据，为简化求解先把题目所给的空间数据通过某种方式投影订正到平面上，然后在水平面上模拟污染分布．

6.2　高度订正方法

第 i 个观测点的污染物浓度指数可表示为

$$I_i = I(x_i, y_i, h_i), \quad i = 1, 2, \cdots, n, \tag{10}$$

其中 x_i, y_i 是观测点的水平位置，h_i 是观测点的海拔高度．将位置函数 f 与海拔高度 h 做变量分离后得到

$$I_i = \widetilde{I}(x_i, y_i) J(h_i) + \varepsilon_i, \tag{11}$$

其中 ε_i 是随机误差，服从正态分布；$\widetilde{I}(x_i, y_i)$ 是浓度随水平位置的变化，$J(h_i)$ 是浓度随高度扩散函数．该式表明第 i 个观测点的污染物浓度可以近似等效为水平扩散和垂直扩散的叠加作用造成．

为得到污染浓度与地形高度之间的关系，必须消除污染指数随水平传播距离变化的影响．选取城区西南角的一处孤立且陡峭的山坡拟合污染传播高度因子函数，该处重金属浓度受高度的影响远远大于受水平距离的影响．

根据比尔-伯格-朗伯定律，假设重金属污染扩散随高度呈指数衰减，于是用指数函数在上述区域对污染传播高度因子函数进行估计，将地累积指数做高度订正，从而将原始样本数据 (x_i, y_i, h_i, I_i) 换算成新的样本 $(x_i, y_i, \widetilde{I}_i)$，从而为接下来的位置估计降维．

6.3 模型求解和污染源的搜索

当只有一个污染源时，(x_i, y_i) 处的浓度分布为

$$\widetilde{I}_i = k\exp\left(-\frac{u_i}{2}\right). \tag{12}$$

当有 N 个污染源时，(x_i, y_i) 处的浓度是 N 个污染源在此处污染效果的叠加

$$\widetilde{I}_i = \sum_{j=1}^{N} k_j \exp\left(-\frac{u_{ij}}{2}\right), \quad u_{ij} = (\mu_i - \mu_0^j)^\top \Sigma_j^{-1}(\mu_i - \mu_0^j),$$

其中 $\mu_i = (x_i, y_i)^\top m$, $\mu_0^j = (x_0^j, y_0^j)^\top$．这里 k_j, Σ_j, $\mu_0^j\,(j = 1, 2, \cdots, N)$ 是待估参数．

基于极大似然估计或最小二乘法[6]进行污染源搜索，即可估计出位置参数 μ_0^j $(j = 1, 2, \cdots, N)$，即为污染源位置．这里污染源的个数可依据AIC模型选择标准或交叉验证法选取．具体的搜索步骤如下：

步骤1 由高度订正公式，将高度为 h_i 的浓度根据指数衰减规律订正到了平面后，可得到订正后的重金属污染等值线图；

步骤2 分析订正后重金属污染物指标的等值线图，找到可能的污染源位置；

步骤3 找出步骤2中观测点内金属污染指标的极大值点，作为污染源位置参数估计的初始位置，搜索区域为污染源周围一定领域；

步骤4 进行区域搜索，根据最小二乘法找出对应的污染源位置；

步骤5 根据AIC准则确定最终的污染源的个数及其位置．

下面以第一类元素(包含Cr和Ni)为例，选取Cr作分析，给出具体的操作过程．根据高度订正后污染分布图4，Cr可能仅有一个单污染源形成或者一个双污染源形成．

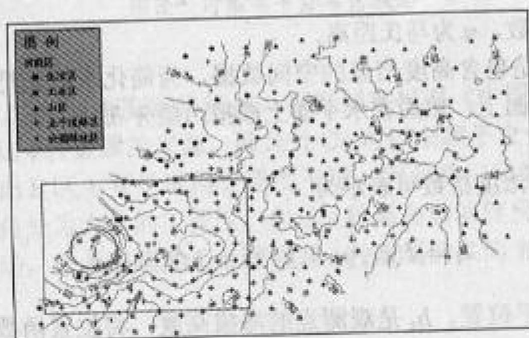


图4: 第一类金属元素的污染分布及污染源搜索

若为单污染源：搜索区域选为 $X \sim [1000, 9000]$, $Y \sim [1000, 9000]$，找到污染源的位置为 $(4592, 4604)$．

若为双污染源：搜索区域选为 $X \sim [1000, 12000]$，$Y \sim [1000, 9000]$，初值取该搜索区域内污染物含量指数的极大值所在位置 $\mu_0\,(4777, 4897)$，$\mu_1\,(10643, 4772)$，计算后找到污染源的位置为 $\mu_0'\,(4580, 4620)$，$\mu_1'\,(15291, 3208)$.

分析上面的搜索结果，可知双污染源得到的两个污染源的位置中，$\mu_1'\,(15291, 3208)$ 位于山区，而且在污染指标等值线分析图中是表现为一个低值，认为这种分析结果并不合理，因此我们认为污染源应为一个单污染源．因此第一类元素的污染是由位于 $(4592, 4604)$ 的单污染源形成的．

7　总结

该模型中所有分析的对象都是重金属浓度的地累积指标，而不是直接使用测到的浓度，因而消除了当地环境背景中金属元素的分布影响；根据污染等级的划分，分析出整个山区都不属于污染区，在污染源考虑时略去右侧的山区，大大简化了问题．测量值是空间数据，包括高程信息，我们在分析污染前对高程先进行订正，以降低模型维数，间接增大样本容量；衰减模型利用比尔吸收定理的衰减原理去拟合样本，得到令人满意的效果；结合污染源扩散污染物浓度的椭球型分布规律，由分布的位置参数确定污染源位置，并且可以处理多个污染源的污染叠加效应．

由于数据信息限制，模型只考虑了点污染源的分布[7]，没有考虑线污染源和其它形式污染源的扩散分布．如果增加信息，则可进一步改进模型：

1)　继续采样，增加样本信息，得到更为真实的重金属污染物浓度指标分布 $C_i(x_i, y_i, h_i)$；

2)　若增加城市的地表信息，如河流、农田等的分布状况，可以估计一些非点源分布的污染源；

3)　考虑风向等引起的污染分布的不对称性，将污染浓度曲面从二元正态模型 $g(u)$ 扩展至偏斜二元正态分布 $2g(u)G(\lambda u)$，其中 $G(\cdot)$ 为二元正态分布函数，λ 为偏斜参数．

参考文献：

[1] 柴世伟. 地累积指数法在土壤重金属污染评价中的应用[J]. 同济大学学报 (自然科学版)，2006, 34(12): 1657-1661
Chai S W. Application of index of geoaccumulation to pollution evaluation of heavy metals in soil[J]. Journal of Tongji University (Natural Science), 2006, 34(12): 1657-1661

[2] 李卫东. 应用多元统计分析[M]. 北京：北京大学出版社，2008
Li W D. Applied Multivariate Statistical Analysis[M]. Beijing: Peking University Press, 2008

[3] 林艳. 基于地统计学与GIS的土壤重金属污染评价与预测[D]. 长沙：中南大学硕士学位论文，2009
Lin Y. Pollution assessement and prediction of heavy metals in soil based on geostatistics and GIS[D]. Changsha: Master Thesis of Central South University, 2009

[4] 徐理超. 阜新市农田土壤重金属污染的空间分析及污染评价[D]. 重庆：西南大学硕士学位论文，2007
Xu L C. Spatial pattern and pollution evaluation of heavy metals in farmland soils of Fuxin city[D]. Chongqing: Master Thesis of Southwest University, 2007

[5] 李月清. 选煤厂产尘环节粉尘颗粒扩散方程的建立[J]. 北京工业职业技术学院学报，2008, 7(1): 84-89
Li Y Q. Construction of diffusion equation in the process of engendering dust in selecting coal factory[J]. Journal of Beijing Polytechnic College, 2008, 7(1): 84-89

[6] 陈杰. Matlab 宝典 (第3版)[M]. 北京：电子工业出版社，2011
Chen J. Matlab Book (3rd Edition)[M]. Beijing: Electronics Industry Press, 2011

[7] 滕葳. 重金属污染对农产品的危害与风险评估[M]. 北京：化学工业出版社，2010
Teng W. Heavy Metal Pollution on Agricultural Products Hazard and Risk Assessment[M]. Beijing: Chemical Industry Press, 2010

Spatial Data Analysis of Heavy Metal Pollution in Urban Soil

SHENG Zhen-feng, HUANG Si, ZHOU Min-qiang
Advisor: CAO Chun-zheng

(Nanjing University of Information Science & Technology, Nanjing 210044)

Abstract: With the rapid development of current industry, heavy metal pollutants (HMP) have made a tremendous impact on the environment. The degree of HMP is measured by the geoaccumulation index (GI), and meanwhile, the degree of HMP in each functional areas is assessed by the pollution load index (PLI). After such analysis, a categorization of heavy metal irons is concluded. Based on the national standard for the categorization of HMP's source and the distributed characteristics of the pollution, the pollution source is inferred for each HMP. Finally, following the revision of the sample's elevation information through the exponential decay model, the pollutant concentration surfaces formed by multi-sources are fitted by the density of weighted mixtures of bivariate normal distributions. The estimates of the location parameters can be regarded as the HMP's sources.

Keywords: heavy metal; pollution analysis; geostatistical analysis; spatial data; geoaccumulation index; mixtures of normal distributions

第28卷 增刊1
2011年11月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 28 Supp.1
Nov. 2011

文章编号:1005-3085(2011)07-0067-08

城市表层土壤重金属污染分析

安传恺[1]， 黄向屹[2]， 连 宸[2]
指导教师： 董子静[2]
(1- 北京大学信息科学技术学院，北京 100871； 2- 北京大学数学科学学院，北京 100871)

编者按: 本文最大的特点是建模过程严谨，行文规范，对问题一和二的讨论尤为突出．作者在使用一种方法时，首先会简要给出其基本原理，讨论该方法的适用性及数据前后处理要注意的问题，得到模型后给出必要的检验．学生严谨的科学精神的培养，正是数学建模竞赛希望达成的最重要的目标．另外，问题二建模的方法比较有特色．

摘 要: 近年来，随着科学技术的发展以及工业化的深入，我国土壤污染对城市环境质量的影响日显突出．随着统计技术方法的发展，土壤污染相关的数据大量增加，如何应用查证获得的海量数据资料开展城市环境质量评价和研究人类活动影响下城市地质环境的演变模式成为人们关注的焦点．问题一给出8种主要重金属元素在该城区的空间分布，并分析该城区内不同区域重金属的污染程度．本文采用在地理统计学中广泛使用的Kriging插值的方法对数据进行拟合，为衡量不同区域重金属污染程度，采用主分量分析评价法获得各个采样点的综合污染指数．问题二要求通过数据分析说明重金属污染的主要原因．本文首先确定城市中土壤重金属污染的七大成因，然后采用层次分析的方法研究它们的权重．

关键词: Kriging插值；主分量分析评价法；层次分析法

分类号: AMS(2000) 30E05; 62J12 **中图分类号:** O241.3; O151.21 **文献标识码:** A

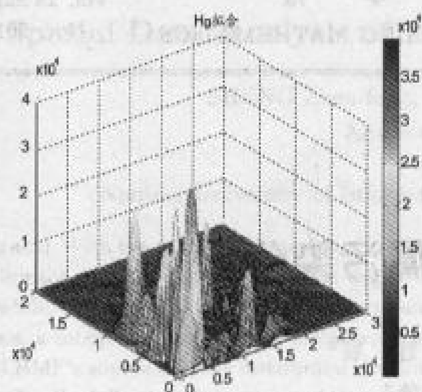
1 问题一(1)模型建立与求解

克里金插值法 题目数据给出了区域内319个采样点表层土壤重金属的浓度．为获得重金属元素在整个城区的空间分布，本文采用克里金[1]二维插值方法处理数据，获得各个重金属元素在城区内空间分布的图像．

克里金[2]插值法借助观测的点$z_i = Z(x_i)$，$i = 1, 2, \cdots, n$处的污染值，通过随机模型计算$Z(x_0)$的最佳线性无偏估计$\hat{Z}(x_0)$，推断区域内$Z(x)$的任意点$Z(x_0)$的污染值．

模型求解 把319个采样点的坐标值作为模型的x_i，8种主要重金属元素在采样点处的浓度及采样点的海拔作为模型中的$z(x_i)$ $(i = 1, 2, \cdots, 319)$，用Matlab编程实现克里金插值法，获得了8种主要重金属元素在该城区的空间分布以及该地区的海拔分布．以Hg为例，见图1和图2，其中1代表生活区，2代表工业区，3代表山区，4代表主干道路区，5代表公园绿地区．

数据观察 克里金插值法对于城市区域中央采样充足的数据拟合得都很好．但是在每一幅*数据图中横坐标5000–8000，纵坐标8000–10000之间有一异常峰．进一步对比原数据表明，该*峰位于原数据最边缘，且周围采样点极少，可能是由采样不足所导致的异常峰．



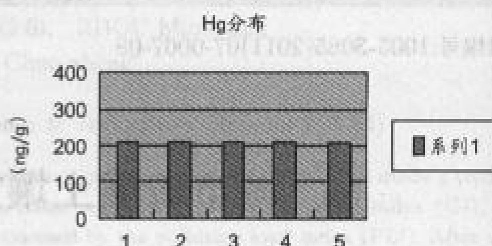


图 1: 重金属Hg在该城区的空间分布

图 2: 重金属Hg在该城区的海拔分布

2 问题一(2)模型建立与求解

主分量分析评价法 题目要求我们分析该城区内不同区域重金属的污染程度．某一区域重金属污染程度由该区域各种重金属的污染情况共同决定，本文用主分量分析评价法得出加权系数，去除了一般赋值方法中的主观因素，因而更为科学．具体步骤如下：

1) 数据标准化处理[2]

由于各参数的量纲与大小不同，首先须将初始变量数值标准化，统一比较的尺度．用下列分指数公式将参数数据标准化

$$x_{ij} = \bar{C}_{ij}/S_j, \quad i=1,2,\cdots,m, \quad j=1,2,\cdots,n,$$

式中 x_{ij} 代表第 i 个样本第 j 种参数的标准化值；$\bar{C}_{ij}$ 是第 i 个样本第 j 种参数的 N 个点位的监测值的平均值；S_j 代表第 j 种参数的最高容许值(背景值范围中最大值)；m 代表样本个数；n 代表参数总数．

2) 主分量分析法[2]

将 m 个样本的 n 个参数数据进行标准化处理，构成一个标准化数据矩阵 $X_{n\times m}$，用Matlab求 X 的协方差阵 $C=\frac{1}{n}XX^T$ 的特征值 $\lambda_i\,(i=1,2,\cdots,n)$ 及相应的特征向量 $\mathbf{v}_i\,(i=1,2,\cdots,n)$ 将特征向量正交化可得正交方阵 V，作变换 $Y=VX$，特征值 λ_i 就是新变量 $\mathbf{y}_i$ 的方差．将 n 个特征值按大小顺序排列 $(i=1,2,\cdots,n)$，相应的 n 个特征向量组成 n 个新变量．方差大的新变量对模型的贡献较大．新变量 $\mathbf{y}_1,\mathbf{y}_2,\cdots$ 分别称为第一主分量，第二主分量，$\cdots$，前面几个主分量构成了样本间的最大变化特征：

$$\begin{cases} y_1 = v_1^{(1)}x_1 + v_2^{(1)}x_2 + \cdots + v_n^{(1)}x_n, \\ y_2 = v_1^{(2)}x_1 + v_2^{(2)}x_2 + \cdots + v_n^{(2)}x_n, \\ \cdots = \cdots + \cdots + \cdots + \cdots, \\ y_n = v_1^{(n)}x_1 + v_2^{(n)}x_2 + \cdots + v_n^{(n)}x_n. \end{cases}$$

前面几个主分量 $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_p\,(p < n)$ 的方差占总体方差的比例为：

$$\rho = \sum_{i=1}^{p} \lambda_i \Big/ \sum_{i=1}^{n} \lambda_i.$$

当 $\rho \geq 0.7$ 时，前 p 个主分量称为公共因子，此时可选用前 p 个主分量代替原来 n 个变量，n 个变量在第 i 个公共因子上的荷载向量为

$$\mathbf{a}_i = \sqrt{\lambda_i}\mathbf{v}^i = \sqrt{\lambda_i}\left(v_1^{(i)}, v_2^{(i)}, \cdots, v_n^{(i)}\right)^T, \quad i = 1, 2, \cdots, p.$$

第 j 个变量在全部 p 个公共因子上荷载的平方和叫做变量的公共性

$$h_j^2 = \sum_{i=1}^{p} a_{ij}^2, \quad j = 1, 2, \cdots, n,$$

从而有

$$h_j = \sqrt{\sum_{i=1}^{p} a_{ij}^2}.$$

它的大小反映了变量 j 在公共性部分的作用或重要性程度．把每个变量公共性的方根 h_j 作为该变量的权重，则构成一个度量环境质量优劣程度的综合指标

$$Y = \sum_{j=1}^{n} h_j x_j,$$

式中 x_j 是用分指数表示的某样本的第 j 种参数的标准化值，它在数据标准化处理时给出；h_j 是计算给出的第 j 种参数的权重．根据 Y 的表达式可以计算出样本环境质量优劣程度的综合指标值，并可按指标值的大小进行环境质量评价．

3）模型实现

本文将8种不同重金属的浓度在319个采样点的数值和背景平均值共320组数据带入模型，将背景值范围最大值作为最大容许值进行标准化处理；再用Matlab实现主分量分析，得到每个采样点和背景值的综合污染指数；最后用背景值的综合污染指数进行归一化处理，得到319个采样点的综合污染指数；再次用Kriging插值法对综合污染指数进行拟合，得到综合污染指数的空间分布和等值线分布如图3和图4，其中横坐标为 x，纵坐标为 y，横坐标一个单位代表300，纵坐标一个单位代表200．

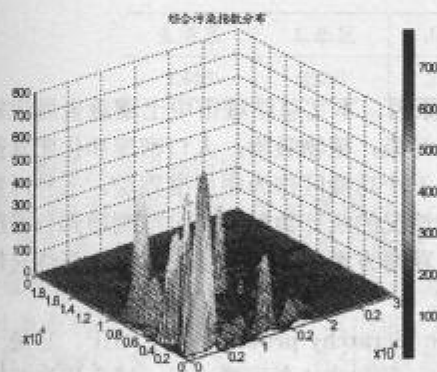


图3：　综合污染指数的空间分布

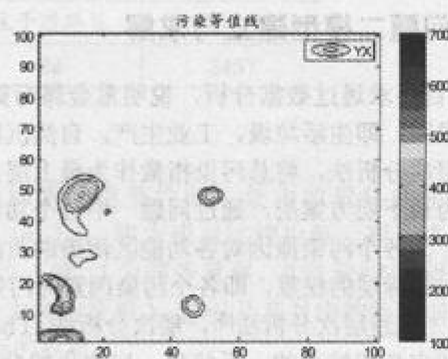


图4：　综合污染指数的等值线分布

对各个功能区污染程度统计，如表1.

将整个城区按横纵坐标等分成5×5的网格，各区域内采样点的综合采样指数做平均后得到各个区域的污染情况统计，见表2，0表示该区域没有采样点.

表1: 各个功能区污染程度统计表

区域＼金属	As (ug/g)	Cd (ug/g)	Cr (ug/g)	Cu (ug/g)	Hg (ug/g)	Ni (ug/g)	Pb (ug/g)	Zn (ug/g)	综合指标
生活区	5.5	295.7	52.8	47.1	208.5	17.1	59.8	195.7	5.0
工业区	5.7	311.9	55.9	60.6	339.0	17.7	64.6	217.8	7.7
山区	5.3	261.8	41.4	34.0	207.0	16.2	52.2	135.8	4.7
主干道路区	5.7	302.9	53.5	55.1	300.5	17.2	61.8	201.5	6.9
公园绿地区	5.6	295.0	52.6	46.1	202.8	17.1	59.1	192.2	4.9

表2: 各个区域的平均污染情况统计表

x＼y	[0, 4000)	[4000, 8000)	[8000, 12000)	[12000, 16000)	[16000, 20000)
[0, 6000)	24.17	7.92	3.86	0.00	0.00
[6000, 12000)	3.11	6.63	4.65	5.50	3.02
[12000, 18000)	20.03	1.37	12.90	2.00	1.44
[18000, 24000)	0.00	1.20	2.80	1.03	1.14
[24000, 30000)	0.00	1.62	0.99	0.97	1.06

由以上两表格，对五种功能区，污染情况从重至轻依次是：工业区，主干道路区，生活区，公园绿地区，山区．在城市内有三块区域污染较为严重，污染源很可能位于其中：$[0,6000]\times[0,4000]$, $[12000,18000]\times[0,4000]$, $[12000,18000]\times[8000,12000]$.

3　问题二模型建立与求解

题目要求通过数据分析，说明重金属污染的主要原因．本文总结城市中土壤重金属污染的七大成因，即生活垃圾，工业生产，自然沉积，交通污染，植物吸附，人口密度和化肥农药．采用层次分析法，将总污染指数作为最上层目标层，五个功能区作为中层准则层，七种污染原因作为最下层方案层，通过问题一中各个功能区污染情况的统计数据，确定准则层到目标层的权重，为各个污染原因对各功能区污染的贡献赋值，获得方案层对准则层的权重，最终获得方案层对目标层的权重，即各个污染因素对污染的影响.

首先概述层次分析法[3]，层次分析法 (The analytic hierarchy process) 简称AHP，它是一种定性和定量相结合的、系统化、层次化的分析方法，一般包含4个步骤：1)　建立层次结构模型；2)　构造成对比较矩阵；3)　计算权向量并做一致性检验；4)　计算组合权向量并做组合一致性检验．下面借助Excel与Matlab进行上述步骤.

1)　建立层次结构模型[3]

对于本问题，目标层的任务是决定七大因素中哪些因素是该城区重金属污染的主要成因．准则层是生活区，工业区，山区，主干道路区，公园绿地区，方案层为生活垃圾，工业生产，自然沉积，交通污染，植物吸附，人口密度和化肥农药七大因素．

通过查阅相关资料[4]，得知可能导致污染的原因如下：

生活垃圾：废弃的生活物品中含有重金属元素，比如电池；工业生产：化学反应产生含重金属的废水废渣；自然沉积：描述没有任何生物作用的情况下重金属在自然界的分布与转化；交通污染：车辆排放含有重金属的尾气；植物吸附：某些植物的根可以吸附土壤中的重金属，形成聚集；人口密度：其他人为活动对重金属含量的影响，如化学实验；化肥农药：化肥农药在土壤中仍可能有残留．

针对题目中的五大类生活区域，它们各自有不同的影响因素，见图 5．生活区：大量生活垃圾，存在交通污染情况，人口密度大．工业区：进行工业生产的区域，工厂附近交通运输频繁，人口密度稍小．山区：山区人口密度较小，存在植物的吸附作用．主干道路区：交通污染比较严重．公园绿地区：交通污染程度轻，植物主要集中在公园绿地，撒播的农药有残留．自然沉积普遍存在于各个区域．

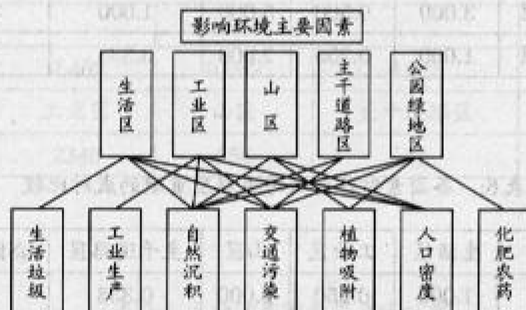


图 5:　五大类区域不同的污染因素

表 3 是准则层与方案层的影响关系 (七大因素依次以 1 到 7 表示)．

表 3:　准则层与方案层的影响关系

生活区	工业区	山区	主干道路区	公园绿地区
1346	2346	356	34	3457

2)　构造成对比较矩阵

比较第 i 个元素与第 j 个元素相对上一层某个因素的重要性时，使用数量化的相对权重 a_{ij} 来描述．设共有 n 个元素参与比较，则 $A=(a_{ij})_{n\times n}$ 称为成对比较矩阵．对于多项指标，量化两两间的相对重要程度，取值为 1–9，1/2–1/9．1 为同等重要，9 表示 i 对 j 极端重要，1/9 表示 i 对 j 极端不重要．求出这一矩阵的最大实特征根 $\lambda_{\max}$ 以及其对应的特征向量，将特征向量的分量和化为 1 后，各分量就是各项指标的权重．

表 4 给出由第一问得到的各区的综合污染指数，转化为各区对污染的贡献的成对比较矩阵，见表 5．根据以上分析得到反映各因素对各区域污染程度贡献的成对比较矩阵，见表 6．

表4: 各区的综合污染指数

生活区	4.993
工业区	7.737
山区	4.730
主干道路区	6.897
公园绿地区	4.866

表5: 各区对污染的贡献的成对比较

	生活区	工业区	山区	主干道路区	公园绿地区
生活区	1.000	0.250	2.000	0.333	1.000
工业区	4.000	1.000	9.000	2.000	4.000
山区	0.500	0.111	1.000	0.200	0.500
主干道路区	3.000	0.500	5.000	1.000	3.000
公园绿地区	1.000	0.250	2.000	0.333	1.000

表6: 各因素对各区域污染程度贡献的成对比较

	生活区	工业区	山区	主干道路区	公园绿地区
生活区	1.000	0.250	2.000	0.333	1.000
工业区	4.000	1.000	9.000	2.000	4.000
山区	0.500	0.111	1.000	0.200	0.500
主干道路区	3.000	0.500	5.000	1.000	3.000
公园绿地区	1.000	0.250	2.000	0.333	1.000
特征值	CI	CR			
5.017	0.004	0.004			
归一权重	0.104	0.462	0.052	0.278	0.104

3) 计算权向量并做一致性检验

对于每一个成对比较矩阵，计算最大特征根及对应的特征向量．定义一致性指标CI、平均随机一致性指标RI，进而相除得到随机一致性比率CR，将其作为检验成对比较矩阵一致性的标准．若矩阵通过检验，特征向量归一化后即为权向量；若不通过，需重新构造成对比较矩阵．检验成对比较矩阵A一致性的步骤如下：

计算衡量一个成对比较矩阵A ($n>1$阶方阵) 一致程度的指标$\mathrm{CI}=\frac{\lambda_{max}(A)-n}{n-1}$，CI非负，且CI越小，一致性越好．

RI称为平均随机一致性指标，它只与矩阵阶数n有关，是固定的经验值，见表7．

计算成对比较矩阵A的随机一致性比率$\mathrm{CR}=\frac{CI}{RI}$．

判断矩阵一致性方法如下：当 $CR < 0.1$ 时，可判定成对比较矩阵 A 具有较好一致性，或其不一致程度在容许范围内；否则必须调整成对比较矩阵 A，直到取得较好一致性为止．表5中各因素权重以及一致性检验情况如表6，从表中可以得知，该矩阵一致性较好．

计算组合权向量并做组合一致性检验：计算最下层各因素对目标层的权重，同理做组合一致性检验．表8是5区污染贡献权重，表9为各种污染因素的最终权重．

表7：平均随机一致性指标与矩阵阶数 n 的关系

矩阵阶数	3	4	5	6	7	8	9
RI	0.51	0.89	1.12	1.25	1.35	1.42	1.46

表8：5区污染贡献权重

		目标层		
0.104	0.462	0.052	0.278	0.104
生活区	工业区	山区	主干道路区	公园绿地区
1346	2346	356	34	3457

表9：各种污染因素的最终权重

生活垃圾	工业生产	自然沉积	交通污染	植物吸附	人口密度	化肥农药
1	2	3	4	5	6	7
0.062	0.340	0.122	0.293	0.037	0.085	0.068

总排序一致性检验公式为

$$\mathrm{CR} = \frac{\sum_{j=1}^{4} a_j CI_j}{\sum_{j=1}^{4} a_j RI_j},$$

即对第二层中采用权重矩阵计算权值的四个区域(生活区，工业区，山区，公园绿地区)的CI, RI进行加权计算．表8中若干个因素分别对四个区域影响，CI来自于反映这些影响的四个矩阵．CR结果为 $0.00922 < 0.1$，故总层次排序一致性较好．

4) 结论

最主要的影响因素是工业生产和交通污染，其他5种污染因素严重性由大至小排序为自然沉积，人口密度，化肥农药，生活垃圾，植物吸附．

4 问题3模型建立与求解(略)

5 问题4模型建立与求解(略)

参考文献:

[1] 百度百科，克里金插值[OL]. http://baike.baidu.com/view/1684023.html, 2011
Baidu Baike,Kriging interpolation[OL]. http://baike.baidu.com/view/1684023.html, 2011

[2] 李祚泳，丁晶，彭荔红. 环境质量评价原理与方法[M]. 北京：化学工业出版社，2004
Li Z Y, Ding J, Peng L H. The Principles and Method of Environmental Quality's Evaluation[M]. Beijing: Chemical Industry Press, 2004

[3] 陈东彦，李冬梅，王树忠. 数学建模[M]. 北京：科学出版社，2007
Chen D Y, Li D M, Wang S Z. Mathematical Models[M]. Beijing: Science Press, 2007

[4] 王济. 贵阳市土壤重金属污染研究[M]. 北京：气象出版社，2011
Wang J. Research on Heavy Metals in Soil of Guiyang[M]. Beijing: Meteorology Press, 2011

Analysis of Heavy Metal Pollution in Urban Topsoil

AN Chuan-kai[1], HUANG Xiang-yi[2], LIAN Chen[2]

Advisor: DONG Zi-jing[2]

(1- School of Electronic Engineering and Computer Science, Peking University, Beijing 100871;
2- School of Mathematical Sciences, Peking University, Beijing 100871)

Abstract: During the rapid pace of the Chinese industrialization in recent years, the environmental side effect, especially the soil pollution, becomes more and more serious. Accumulating many data of the soil pollution, researchers pay increasing attention on unveiling the relationship between human activities and the environment by using statistical and machine learning tools. This paper evaluates the pollution distribution and analyzes the pollution level of eight major heavy metals in a certain urban area by utilizing Kriging interpolation and principal component analysis in part 1. In the hierarchic analytic way, in part 2 the paper finds out the main reasons (factors) underlying soil pollution and substantiates the relative weights of these factors.

Keywords: Kriging interpolation; principal component analysis; analytic hierarchy process

第28卷 增刊1　　工 程 数 学 学 报　　Vol. 28 Supp.1
2011年11月　　CHINESE JOURNAL OF ENGINEERING MATHEMATICS　　Nov. 2011

文章编号:1005-3085(2011)07-0075-08

土壤表层重金属污染研究

王九亚， 李达超， 朱怀槿
(中山大学数学与计算科学学院，广州 510275)

编者按：本文研究了向外排放重金属污染物的恒定污染源和废弃污染源的区分问题，结合重金属浓度局部最大值处地形曲面和浓度曲面的梯度夹角给出判定准则，并进行仿真实验．这是一个很有意义的问题，值得读者对此进行进一步的深入研究．

摘 要：本文给出了一种确定重金属污染源位置的有效方法并尝试对污染源进行分类．首先，我们通过分析实际数据及仿真实验，得到土壤表层重金属污染的传播模式：污染物倾向于向低海拔传播；设定了不同初值的扩散仿真实验呈现出两种不同类型的污染传播模式，即恒定污染源和废弃污染源．结合得到的规律，我们根据重金属浓度局部最大值处地形曲面和浓度曲面的梯度场夹角变化趋势，筛选出真正的污染源；并且使用Gauss分布和指数分布两种扩散分布拟合出局部重金属浓度值，并通过最小二乘法比较误差进行污染源类型的确定．

关键词：重金属污染；梯度场夹角；扩散模型；恒定污染源；废弃污染源

分类号：AMS(2000) 90C05　　中图分类号：O29　　文献标识码：A

1 基本假设

1) 本模型只讨论土壤表层的污染，不考虑污染浓度随着土壤深度的变化和分布情况，即假设重金属只在土壤表面传播；

2) 重金属背景值代表了该地区土壤中各重金属元素浓度的本底值，实际测量值则在本底值的基础上反映人为活动造成的污染程度；

3) 重金属污染源总是重金属浓度的极大值点，个别污染最严重的点可作为主要污染源．

2 建立模型

2.1 模型结构

模型结构如图1所示．

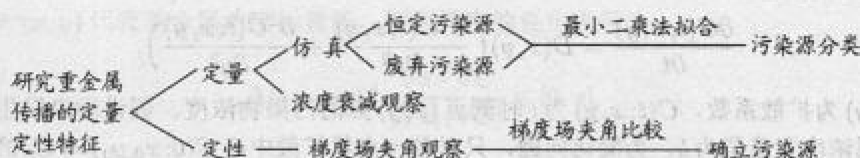


图 1: 分析步骤

2.2 模型的确立

2.2.1 重金属污染分布

1) 重金属浓度分布

城市的表层是不规则曲面，而重金属在表层迁移，考虑不同测量点方位和位置上的实际距离，不应直接在平面上对浓度进行插值：

a) 利用实测点数据 $P_i(x_i, y_i, h_i)$ 在平面网格上对海拔值进行插值，得到城区曲面地形图；

b) 根据上一步中得到的网格曲面和重金属浓度实测值 C_i 进行曲面上的浓度插值，得到重金属污染在曲面上的分布函数.

第一步中采用了Matlab 4格点样条插值方法，此方法计算得出的曲面较为光滑，效果较好. 第二步中采用了Kriging插值法进行污染浓度插值[1].

2) 污染级别评价

采用地累积指数法[2]对重金属污染程度进行评级. 地累积指数法是通过计算实测浓度值 C 与当地背景值 Ba 之间的关系来评价人类活动造成的污染严重程度. 其计算式如下

$$I_{\text{geo}} = \frac{\log_2 C}{k \times Ba}, \tag{1}$$

其中 k 是为考虑各地岩石差异可能会引起背景值的变动而取的系数，一般取值为1.5；I_{geo} 为地累积指数. 根据 I_{geo} 将污染等级分为6级[2]，如表1所示.

表1: 地累积指数与污染程度分级

I_{geo}	<0	0–1	1–2	2–3	3–4	4–5	>5
级数	0	1	2	3	4	5	6
污染程度	无	无–中	中	中–强	强	强–极强	极强

2.2.2 重金属污染传播特征分析

1) 衰减分布函数的特征(定量)

首先，确定一个所在地势较为平坦的严重污染点，即主要污染源，设为原点 O，这样就可以排除地形对衰减的影响. 以 O 为中心，划定污染级别高于 I_i 的区域 D，考察在 D 上的污染衰减分布 $d = d(r)$，其中 r 为 D 上任一点与 O 的距离，寻找相似的分布，以便利用这些已知的分布函数去拟合 $d(r)$.

2) 基于扩散模型的仿真实验(验证定量特征分析)

由于土壤中的重金属传播极类似于固体与固体间的扩散模型，故尝试使用扩散方程去描述这一过程并进行仿真试验，考察污染源周围空间中分布的衰减随时间变化的过程.

Fick's扩散定律[3]：扩散物质在单位时间内沿法线方向流过曲面单位面积的质量与物质浓度沿该方向的导数成正比

$$\frac{\partial C(t;x,y)}{\partial t} = D(x,y)\left(\frac{\partial^2 C(t;x,y)}{\partial x^2} + \frac{\partial^2 C(t;x,y)}{\partial y^2}\right), \tag{2}$$

其中 $D(x,y)$ 为扩散系数，$C(t;x,y)$ 为 t 时刻点 (x,y) 点的污染物浓度. 假设污染发生前，该地区的重金属浓度背景值为 b. 为简化问题，只对单一点状扩散中心 $C(0;x_0,y_0) = C_0$ 的扩散系统进行时间和空间离散化的仿真

$$C(0;x,y) = \begin{cases} C_0, & x = x_0 \wedge y = y_0, \\ b, & \text{其它}. \end{cases} \tag{3}$$

考虑到污染系统的封闭性，我们对其边界条件进行以下两种假设：

i) 边界条件恒定

$$C(t, x_0, y_0) = C_0. \tag{4}$$

ii) 无边界条件

另外，我们称密度函数符合下列函数的分布为**指数分布**

$$f(x, y; \lambda) = \lambda^2 e^{-\lambda\sqrt{x^2+y^2}}, \tag{5}$$

其中$\lambda > 0$是指数分布的率参数．类似地，Gauss **分布**的密度函数为

$$f(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{6}$$

其中σ是Gauss分布的标准差．

从仿真所得结果可以看出，在不同的边界条件下衰减呈现不同的污染传播模式．当污染源不断向外排放新的污染物，即中心污染浓度不变的情况下，污染衰减随径向呈现指数分布，此时称污染源为为**恒定污染源**；而当污染源不向外排放新的污染物，即整个系统内污染物总量不变时，污染物从中心向外扩散呈现Gauss分布的衰减模式，此时称污染源为**废弃污染源**．

3) 衰减分布函数的特征(定性)

由于水流、大气、泥沙等因素是自然中影响重金属污染传播的重要因素，而这些因素往往很容易随着地形运动，例如，水总是往低处流，而泥沙总是往山底滑，因此地形对重金属污染的传播有着重要的作用．

我们使用梯度来描述地形的影响．梯度代表向量场中方向导数最大的方向，地形梯度即海拔上升最快的方向，浓度梯度即污染上升最快的地方．考虑处于斜坡上的污染源(处于地形曲面极小值和极大值处的概率可近似看作为0)，其浓度和地形曲面的梯度场夹角较非污染源有明显区别：污染源点下方的斜坡，两向量场梯度夹角为锐角，而污染点上方的斜坡，两向量场梯度的夹角为钝角．相反，若不存在污染源，各点梯度夹角始终接近180°．模型的求解中得到的图形将会展现出这一定性关系是十分显著的．

2.2.3 基于梯度场夹角比较法确定污染源

我们要利用上述得到的传播特征来筛选出真正的污染源．除了污染源本身之外，一些地形低洼处，也容易汇集从别处传播来的重金属，因此需要在有一定污染程度的重金属浓度极大值点中辨别出哪些是真正的污染源，而哪些仅仅是地形造成的重金属汇集点．

设空间函数$z = z(x, y)$，则标量场$z(x, y)$的梯度即为

$$\nabla z = \Big(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\Big). \tag{7}$$

设$C = C(x, y)$代表重金属浓度标量场，则两梯度的夹角表示为

$$\Delta\theta = \cos^{-1}\frac{\nabla z \cdot \nabla C}{|\nabla z|\,|\nabla C|} \in [0, \pi]. \tag{8}$$

先选取污染度高于I_t的污染区域，再在区域内选择金属浓度极大值点，称其为热点．作$\Delta\theta$在地形上的分布，观察热点附近夹角的大小．

判断的标准：

夹角的大小：若一个热点为污染源，其周围的点中就会存在两梯度夹角较小的情况；

夹角的分布：若一个热点为污染源，则在地形梯度的一面，两梯度场夹角较小，而在另一侧，夹角较大．故污染源的另一个重要特征是，存在夹角剧烈变化的分界点或分界线．

2.2.4 基于最小二乘法的污染源分类

由扩散仿真试验可知，恒定污染源的浓度衰减分布近似满足指数分布，而废弃污染源的浓度衰减分布则近似满足Gauss分布．因此可以在指数分布和Gauss分布的密度函数构成的空间内，寻找与污染源周围数据最小二乘距离最小的曲面，从而判定其污染类型．为了在最小二乘中对数据进行归一化，定义纵向的拉伸系数k和平移系数b．由此得到新的用于拟合的径向指数曲面和径向高斯曲面表达式

$$f(x,y;\lambda,k_1,b_1)=k_1\lambda^2e^{-\lambda\sqrt{x^2+y^2}}+b_1, \tag{9}$$

$$f(x,y;\sigma,k_2,b_2)=\frac{k_2}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}+b_2. \tag{10}$$

取出污染源P_i附近，污染级别高于I_i的区域D中重金属浓度散点值$\{(x_j,y_j,c_j)\}_{j=1}^{n}$；取定$k$, b的适当范围后．原问题转化为两个规划问题

$$\min_{\lambda,k_1,b_1}\quad err_e=\sqrt{\frac{\sum\limits_{i=1}^{n}(f(x_i,y_i;\lambda,k_1,b_1)-c_i)^2}{n}}, \tag{11}$$

$$\min_{\sigma,k_2,b_2}\quad err_g=\sqrt{\frac{\sum\limits_{i=1}^{n}(f(x_i,y_i;\sigma,k_2,b_2)-c_i)^2}{n}}. \tag{12}$$

若$err_e<err_g$，则此污染源很可能是恒定污染源；若$err_e>err_g$，则此污染源更有可能为废弃污染源．

3 模型求解

3.1 重金属污染分布

根据所提供的该市各测量点的空间方位和重金属浓度的数据信息，插值计算得到该市的地累积污染指数分布图，如图2所示，以Hg为例．

3.2 重金属污染源分析

3.2.1 分析污染的传播特征

1) 观察浓度分布衰减图

以Hg为例，我们找到并放大污染最为严重的个别点，观察它们附近的污染浓度分布情况．由图3可知，代表点1周围的衰减模式与指数分布近似，而代表点2周围的浓度衰减则较符合Gauss正态分布的模式．为验证观察结果，并进一步定量分析这些分布的特征，我们使用扩散实验仿真模拟重金属污染的过程．

2)仿真实验

将图4与图5两张图进行比对，可以看出，恒定污染源仿真得到的污染分布与径向的指数分布近似，而废弃型污染源仿真结果较符合高斯分布．

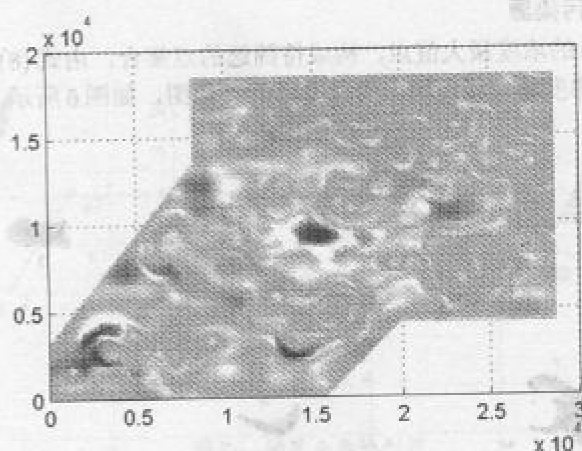
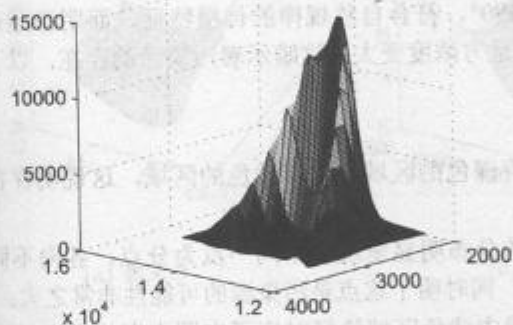
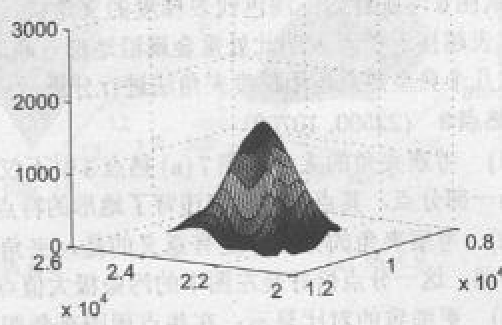


图 2:　地累积污染指数分布图



(a) 代表点 1 (13800, 2400)



(b) 代表点 2 (22400, 10500)

图 3:　污染深度分布情况

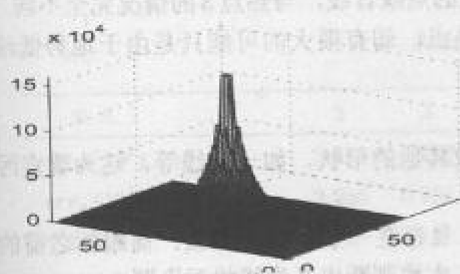


图 4:　恒定污染源

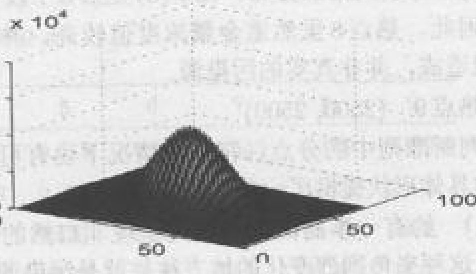


图 5:　废弃污染源

3.2.2 筛选潜在的污染源

取所有污染度大于3的浓度极大值点，构成待筛选的点集合. 由式(8)，利用Matlab绘出金属Hg污染浓度分布的梯度场与城市地形的梯度场间夹角图，如图6所示.

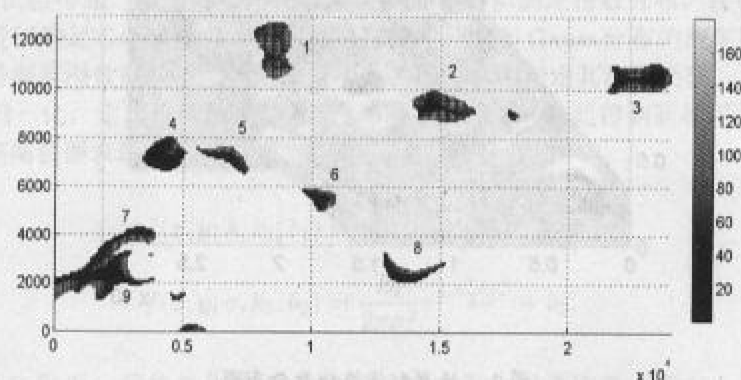


图6: Hg污染浓度梯度场与地形梯度场夹角图

从图6可以看到：浅色代表梯度的夹角接近180°，符合自然规律的传播特征，而阴影越深则代表越接近0°，表明此处重金属沿地形升高的地方浓度变大，这暗示着污染源的存在. 以下将取几个典型热点运用梯度夹角法进行分析示例.

热点3 (22500, 10700)

1) 考察夹角的大小：图7(a)热点3中不仅有深色的区域，也有浅色的区域，这说明存在相当一部分点，其点浓度梯度违背了地形的特点；

2) 考察夹角的分布：更有意义的是，夹角的分布明显呈现以中间一点为分点，各向不同的格局，这一分点恰好是左图中的污染极大值点，同时明了这点是污染源的可能性非常之大；

3) 更细致的对比显示：在热点周围夹角图像中浅色区域恰好对应了左图中海拔升高的区域，而海拔减小的区域则显示出了异常的深色角度值，验证了我们对于梯度夹角的分析.

因此，热点3是一个真正的污染源.

热点8 (8800, 12300)

1) 考察夹角的大小：整片区域中梯度的夹角几乎完全大于100°，说明此区域金属浓度近似按照自然沉降的趋势分布，这恰好被浓度图中高浓度区域完全积聚在洼地印证；

2) 考察夹角的分布：图7(b)热点8中没有突变的点或者线，与热点3的情况完全不同.

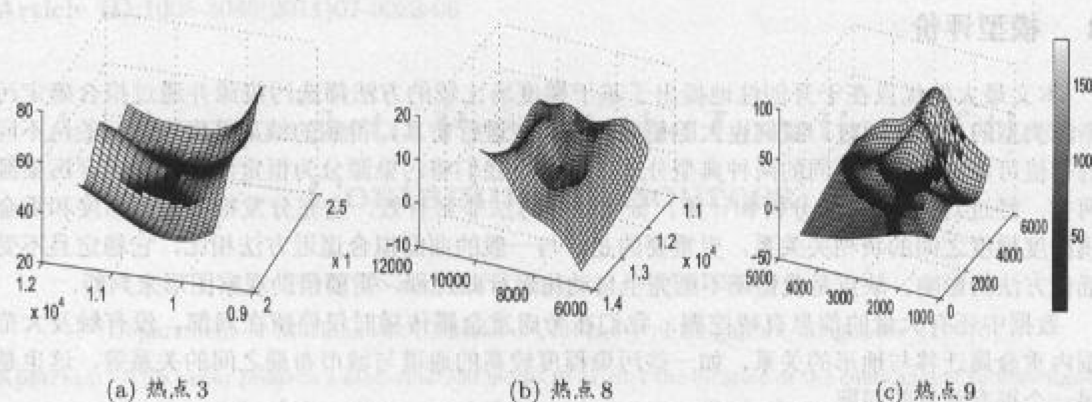
因此，热点8虽然重金属浓度值较高，并形成热点，却有很大的可能只是由于地势低洼而积聚造成，并非真实的污染源.

热点9 (2500, 2500)

判断准则中的分点，在一些情况下也有可能形成其他的形状，如一条线等，这为确定污染源的具体形状提供了一些线索.

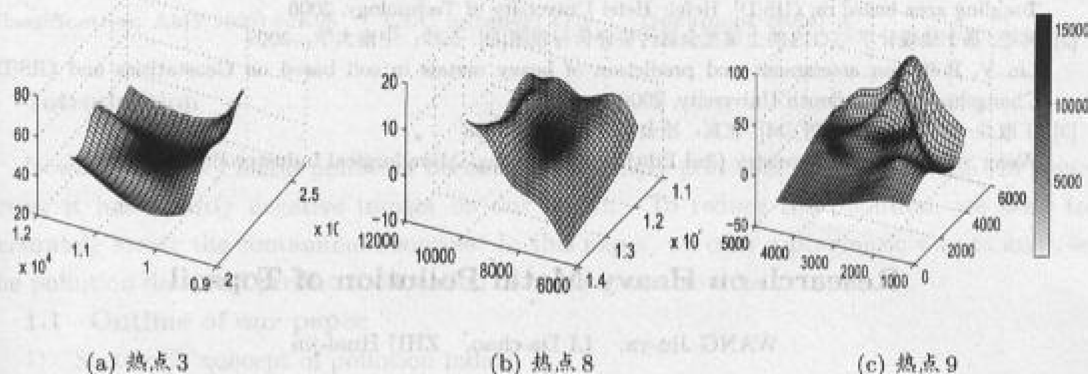
1) 约有一半的面积上夹角呈现非自然的特征，且存在一条明显的界线. 而根据之前的推断，这种夹角剧烈变化的地方往往就是污染源所在，大致判断出其为线状污染源；

2) 结合图8，图7中最热的“C”形热圈的边界在左图中恰好位于半山腰上. 故判定，热点9处存在线状污染源.



(a) 热点3　(b) 热点8　(c) 热点9

图7: 梯度夹角地形图



(a) 热点3　(b) 热点8　(c) 热点9

图8: 深度地形图

3.2.3 污染源分类

下面是以Hg的污染分布为例，使用最小二乘拟合以及梯度夹角比较法法判定出的污染源类型，如表2所示.

表2: 污染源分类

热点	1	2	3	4	5	6	7	8	9
$err_e\,(10^3)$		2.315	0.197	0.208	0.318	0.098	2.543	3.389	2.980
$err_g\,(10^3)$		2.285	0.208	0.133	0.375	0.114	3.104	3.928	3.477
污染源类型	非污染源	废弃	恒定	废弃	恒定	恒定	恒定	恒定	恒定
污染源形状	非污染源	点状	点状	点状	线状	点状	线状	线状	线状

4 模型评价

本文最大的优点在于开创性地提出了基于梯度场比较的方法筛选污染源并通过拟合确定污染源类型的方法．同时，我们也大胆使用扩散模型进行仿真，得到的结果是出人意料的，不同的初值可以造成完全不同的两种典型分布，这引发我们将污染源分为恒定污染源和废弃污染源两类．经过对Hg金属的分析和计算，证明这种方法非常有效．它充分发掘了地形梯度和重金属浓度梯度之间的负相关关系，更重要的是，与一般的曲面拟合逼近方法相比，它稳定且不受插值方法的影响．缺点是我们还不能完全自动化运行此过程，需要借助观察图形来判断．

数据中还有大量的信息有待挖掘．我们在考虑重金属传播时仅停留在局部，没有触及大范围内重金属迁移与地形的关系，如一些污染程度较高的通道与城市布局之间的关系等，这也是另一个很有意思的问题．

参考文献：

[1] 祁轶宏．基于GIS的铜陵地区土壤重金属的空间分布及污染评价[D]．合肥：合肥工业大学，2006
Qi Y H. The study on the spatial distribution and pollution assessment of heavy metal elements in soil of Tongling area based on GIS[D]. Hefei: Hefei University of Technology, 2006

[2] 林艳．基于地统计学与GIS的土壤重金属污染评价与预测[D]．长沙：中南大学，2009
Lin Y. Pollution assessment and prediction of heavy metals in soil based on Geostatitics and GIS[D]. Changsha: Central South University, 2009

[3] 王淑兰．物理化学(第3版)[M]．北京：冶金工业出版社，2008
Wang S L. Physical Chemistry (3rd Edition)[M]. Beijing: Metallurgical Industry Press, 2008

Research on Heavy Metal Pollution of Topsoil

WANG Jiu-ya, LI Da-chao, ZHU Huai-jin

(School of Mathematics & Computational Science, Sun Yat-sen University, Guangzhou 510275)

Abstract: An effective method for locating the pollutant source and a categorization of pollutant sources are proposed. We first obtain spread pattern of heavy metals in topsoil through analysis of real data and simulation results. With different initial values, we discover two different types of pollutant sources and the corresponding spread patterns. One is constant pollutant source and another is abandoned pollutant source. Then we locate real pollutant source by comparing the angle between gradients of floor surface and density surface. Finally, we categorize these pollutant sources by using surface fitting with Gaussian distribution and exponential distribution based on the least square technique.

Keywords: heavy metal pollution; angle between gradients; diffusion model; constant pollutant resource; abandoned pollutant resource

Nov. 2011 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Vol. 28 Supp. 1

Article ID:1005-3085(2011)07-0083-08

An Approach to Predict the Location of the Contaminant Sources

LI Ze-zhe, XU Jing-wei, XU Wen-ji

(Department of Mathematics, National University of Singapore, Singapore 119076)

Abstract: This paper presents a mathematical model to predict the location of the contaminant sources based on the profile of the pollution level of the against region. To do the prediction, we assume certain sampling sites to be the sources and simulate the pollution level of the region. Numerous mathematical tools are used to simulate the diffusion of heavy metal through soil and air. By varying the assumption of the sources and comparing the simulated situation under different assumptions with the real situation, we obtain the optimal number and location of the sources. Computer implementation of the model gives a reasonable result.

Keywords: contaminant sources; location; heavy metal pollution

Classification: AMS(2000) 65K05 **CLC number:** O29 **Document code:** A

1 Introduction

Nowadays, heavy metal pollution becomes increasingly prevalent in our daily life. In some areas, it has a fairly negative impact on our health. To reduce the pollution, we need to accurately locate the contaminant sources. In this paper, we offer a pragmatic way to analyse the pollution data and predict the location of contaminant sources.

1.1 Outline of our paper

1) Section 2: concept of pollution index;

2) Section 3: theoretical fundament of our model;

3) Section 4 and 5: results, conclusion, assessment and generalisation of our method.

1.2 General assumptions

1) The density of heavy metal of the sample site is the average density of the square kilometre;

2) The land shape can be approximated by the smooth surface connecting all the sample sites;

3) Only diffusion through soil and diffusion through air are considered in our model. Other types of propagation may exist, but are not statistically significant;

4) The source of contaminant is located at one of the sample sites.

2 Pollution index

To measure the level of heavy metal pollution for each site, we adopt the Nemerow pollution index[1].

Single contamination index $P_i = C_i/S_i$, where P_i is the single contamination index of heavy metal pollutant i, C_i is its real concentration, and S_i is its regional background value.

Nemerow pollution index[2]

$$P_c = \left[(P_{\max}^2 + P_{\text{ave}}^2)/2\right]^{\frac{1}{2}},$$

where P_c is the Nemerow pollution index of the site, $P_{\max}$ is the highest single contamination index among different heavy metals, and P_{ave} is the average single contamination index.

3 Mathematical model to predict the location of contaminants

In this section, we present the theoretical work of our model.

We will first assume one of the sites in the region to be the contaminant source and simulate the pollution level of the entire region under this assumption. Normal distribution is used to approximate the distribution of heavy metals in a flat region. Several correction factors, such as the altitude difference, the smoothness of land path, the direction and strength of wind, are taken into consideration. Then, we compare the simulated pollution level of the region with the real situation provided in the excel file to test the degree of similarity (using mean square error). We then change our assumption about the contaminant sources and do the simulation. After all possible assumptions are simulated; the assumption with the greatest degree of similarity is concluded to be the location of the contaminant sources.

3.1 Sub-model one: diffusion through soil

As the real world operates, normal distribution is the most general objective distribution law. Thus, it is reasonable to use normal density function to approximate the distribution of heavy metal in a flat region

$$f_A(B) = \frac{1}{\sqrt{2\pi}\sigma_B} \exp\left(-\frac{d_{A,B}^2}{2\sigma_B^2}\right),$$

where A is the assumed contaminant source, B is the point whose level of pollution is to be simulated under the assumption that A is the contaminant source, $d_{A,B}$ is the distance between A and B, and $f_A(B)$ denotes the pollution concentration at B.

Since the region under investigation is far from a flat region, we need to take into account the relief of the terrain between A and B. Thus, we introduce two modifications to the above formula: land smoothness and difference in altitude.

Since the rougher the land path between A and B, the more difficult for heavy metal to propagate, we use the correlation coefficient r to measure the smoothness of the land path and modify the formula as follows

$$F_A(B) = |r| \times f_A(d_{A,B}).$$

The correlation coefficient is calculated by using the point B, the contaminant source point A and all other site points whose x, y coordinates lie near (<1 km) the line connecting A and B:

$$r = \frac{\sum_{i=1}^{n}(h_i - \overline{h})(d_i - \overline{h})}{\sqrt{\sum_{i=1}^{n}(h_i - \overline{h})^2 \sum_{i=1}^{n}(d_i - \overline{d})^2}}.$$

Here we assume there're n site points lying near the line AB. Then we get the projection of these points on the vertical plane containing A and B, and denote them as $x_1, x_2, \cdots, x_n$. For $x_i\,(i = 1, 2, \cdots, n)$, h_i is its altitude, and d_i is the horizontal distance between this point and A.

Besides the above modification, we also need to consider the difference in altitudes of A and B. Here, we need to use slop k and the oblique angle θ:

$$k = \frac{h_B - h_A}{d_{A,B}}, \quad \theta = \arctan(k).$$

Since we just compare the final difference between the simulated pollution and real pollution to predict the location of the contaminant source, we can just use an assumed function that has the same properties of the real distribution function to calculate.

As far as it's concerned, when the terrain between A and B is a straight line and A is above $B\,(\theta < 0)$, the steeper the slope is, the wider the pollutants are able to spread. Accordingly, the larger θ is, the greater σ_B will be. Besides, when $(\theta < 0)$, $\sigma_B > \sigma$. The analysis of situation when A is lower than $B\,(\theta \geq 0)$ is similar. What's more, when $(\theta = 0)$, $\sigma_B = \sigma$ according to the definition; while when θ goes to infinity, the pollution will not spread upward, and thus $\sigma_B = 0$. Then we can modify σ_B by θ:

$$\sigma_B = \sigma \cos\theta, \quad \theta \geq 0, \quad \sigma_B = \frac{\sigma}{\cos\theta}, \quad \theta < 0.$$

Here σ is the standard deviation of the normal distribution where the land is plain.

The σ_B function we define obeys all those properties mentioned above. So it is reasonable and well-defined.

Therefore, we obtain the final distribution function of the heavy metal pollution propagated via soil:

$$F_A(B) = |r| \times \frac{1}{\sqrt{2\pi}\sigma_B} \exp -\frac{d_{A,B}^2}{2\sigma_B^2}.$$

3.2 Sub-model two: diffusion through air

In this sub-model, we still assume that the distribution of pollution level can be approximated by a normal distribution centered at the contaminant source. However, we need to take into consideration some other influencing factors: wind strength and wind direction, which may change the shape of the distribution.

First, assume that the direction of wind is completely random, and then the distributions of the concentrations of heavy metals will follow normal distribution, which centres at the original point (contaminant source), and thus $x \sim N(0, \delta^2)$.

Then based on the analysis of the spatial distribution of pollution provided in the excel files, we find out that the distribution of heavy metals is severely skewed towards the left-bottom corner of the graph. Thus, it is reasonable for us to assume that the direction of wind follows the direction of the vector $\langle -1, -1 \rangle$. We assume that the strength of wind is random over a long period. Thus, the wind will shift the normal distribution curve. Hence the mean of the final normal distribution is approximately the uniformly distribution: $\mu \sim U(0, L_{\text{average_strength}})$, where $L_{\text{average_strength}}$ is the distance that the wind with average strength

can blow the substances containing heavy metals forward in the direction of the wind. Thus, we can calculate the distribution along the direction $\langle -1, -1\rangle$ by stochastic analysis

$$f(x\,|\,\mu) = \frac{1}{\sqrt{2\pi}\delta}\exp - \frac{(x-\mu)^2}{2\delta^2}, \quad f(\mu) = \frac{1}{L_{\text{average_strength}}}.$$

Then

$$f(x) = \int_0^{L_{\text{average_strength}}} f(x,\mu)\mathrm{d}\mu = \frac{\sqrt{\pi}\left(\operatorname{Erf}\left[\frac{L_{\text{average_strength}}-x}{\sqrt{2}\delta}\right] + \operatorname{Erf}\left[\frac{x}{\sqrt{2}\delta}\right]\right)}{2\sqrt{\pi}\delta L_{\text{average_strength}}},$$

where

$$\operatorname{Erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t.$$

Then we need to know how it will change the mean and variance of the normal distribution mean and variance which show how the wind influences the location and spread of the pollution. We attain this aim in the fallming way:

$$E(X\,|\,\mu=\mu_0) = E\left(N(\mu=\mu_0,\delta^2)\,|\,\mu=\mu_0\right) = \mu_0.$$

Hence

$$E(X) = E\left(E(X\,|\,\mu)\right) = E\left(E\left(N(\mu,\delta^2)\,|\,\mu\right)\right) = E(\mu) = \frac{L_{\text{average_strength}}}{2}.$$

Then

$$\operatorname{Var}(X) = E\left(\operatorname{Var}(X\,|\,\mu)\right) + \operatorname{Var}\left(E(X\,|\,\mu)\right),$$
$$\operatorname{Var}(X\,|\,\mu=\mu_0) = \operatorname{Var}\left(N(\mu=\mu_0,\delta^2)\,|\,\mu=\mu_0\right) = \delta^2.$$

Thus we have

$$\operatorname{Var}(X) = E\left(\operatorname{Var}(X\,|\,\mu)\right) + \operatorname{Var}\left(E(X\,|\,\mu)\right) = E(\delta^2) + \operatorname{Var}(\mu) = \delta^2 + \frac{L_{\text{average_strength}}}{12}.$$

The above analysis is based on two points which lie exactly in the same direction as the wind vector, namely $\boldsymbol{e}$. Then we will generalize the aforementioned case to any two points. Suppose that the contaminant point is $A(x_a, y_a)$, another point is $B(x_b, y_b)$, and then there are two vectors, namely: $\boldsymbol{e} = \langle -1, -1\rangle$; $\boldsymbol{AB} = \langle x_b - x_a, y_b - y_a\rangle$.

We can then calculate the angle between these two vectors:

$$\cos\alpha = \frac{(\boldsymbol{AB}\cdot\boldsymbol{e})}{(|\boldsymbol{AB}||\boldsymbol{e}|)}.$$

Knowing the angle between these two vectors, we can obtain a new random variable which suits the general cases by changing the uniform distribution $\mu \sim U\left(0, \cos\alpha L_{\text{average_strength}}\right)$ and we plug it into the previous equations in the following way:

$$f(x) = \frac{\sqrt{\pi}\delta\left(\operatorname{Erf}\left[\frac{L_{\text{average_strength}}-x}{\sqrt{2}\delta}\right] + \operatorname{Erf}\left[\frac{x}{\sqrt{2}\delta}\right]\right)}{2L_{\text{average_strength}}\cos\alpha\sqrt{\pi\delta^2}}, \quad E(X) = \frac{L_{\text{average_strength}}}{2}\cos\alpha,$$

$$\operatorname{Var}(X) = E\left(\operatorname{Var}(X\,|\,\mu)\right) + \operatorname{Var}\left(E(X\,|\,\mu)\right) = E(\delta^2) + \operatorname{Var}(\mu) = \delta^2 + \frac{L_{\text{average_strength}}}{12}\cos\alpha.$$

By using this method, we can calculate the concentration of the ending point by knowing the contaminant point given that it diffuses only through air.

3.3 Combination of the sub-models

From the above two sub-models, we obtain the distribution of the heavy metal pollution propagated via soil and the distribution of the pollution diffused via air. Now we combine these two distributions to get the final distribution of the pollution.

To calculate the final distribution, we assume a fixed proportion τ of the pollution is propagated via soil. Accordingly, $1-\tau$ of the pollution is diffused by air.

Assuming A is the source, we want to get the concentration of the pollution at B. Firstly, we need to obtain the ratio of the pollution concentration at B to the concentration at A diffused via soil and air respectively by the two sub-model. We use C_{A1}, C_{A2}, C_{B1}, C_{B2} to denote the pollution concentrations diffused via soil and air at A and B, respectively.

According to the sub-models, we have

$$\frac{C_{B1}}{C_{A1}}=\frac{r\times\frac{1}{\sqrt{2\pi}\sigma_B}e^{-\frac{d_{A,B}^2}{2\sigma_B^2}}}{\frac{1}{\sqrt{2\pi}\sigma}}=r\times\frac{\sigma}{\sigma_B}e^{-\frac{d_{A,B}^2}{2\sigma_B^2}},$$

$$\frac{C_{B2}}{C_{A2}}=\frac{\frac{\left(\sqrt{\pi}\delta\left(\mathrm{Erf}\left[\frac{L_{\mathrm{average_strength}}-d_{A,B}}{\sqrt{2}\delta}\right]+\mathrm{Erf}\left[\frac{x}{\sqrt{2}\delta}\right]\right)\right)}{2L_{\mathrm{average_strength}}\cos\alpha\sqrt{\pi\delta^2}}}{\frac{\sqrt{\pi}\delta\,\mathrm{Erf}\left[\frac{L_{\mathrm{average_strength}}}{\sqrt{2}\delta}\right]}{2L_{\mathrm{average_strength}}\sqrt{\pi\delta^2}}}=\frac{\left(\mathrm{Erf}\left[\frac{L_{\mathrm{average_strength}}-d_{A,B}}{\sqrt{2}\delta}\right]+\mathrm{Erf}\left[\frac{x}{\sqrt{2}\delta}\right]\right)}{\cos\alpha\,\mathrm{Erf}\left[\frac{L_{\mathrm{average_strength}}}{\sqrt{2}\delta}\right]}.$$

Then, the final ratio of the total pollution concentration at B given A is the contaminant source

$$\rho=\frac{C_B^A}{C_A}=\tau\frac{C_{B1}}{C_{A1}}+(1-\tau)\frac{C_{B2}}{C_{A2}}.$$

Hence, the estimated pollution concentration at B is $C_B^A=\rho\times C_A$, where C_A is the real pollution concentration we collect at A.

In our model, we use Nemerow pollution index to represent the pollution. Hence, the formula will be $P_B^A=\rho\times P_A$, where P_B^A is the estimated Nemerow pollution index at B, given A is the contaminant source, and P_A is the real Nemerow pollution index at A.

By implementing the above method, we can get one dataset of the estimated Nemerow pollution indexes of pollution under the assumption that a certain point is the contaminant source. For instance, we assume point i is contaminant source, and then we get P_k^i which denotes the estimated Nemerow pollution index of pollution at point k given i is the source. So we can get 319 such P_k^is.

To compare the estimated value with the real one, we use the following formula

$$\mathrm{MSE}_i=\frac{\sum_{k=1}^{319}\left(P_k^i-P_k\right)^2}{319},$$

where P_k is the real pollution concentration at point k.

Applying the same method to every point (changing i we mentioned in the last paragraph), we can get 319 datasets of the estimated values, and we also get 319 MSE_is.

If there's only one contaminant source, we shall find out $\min_i\{\mathrm{MSE}_i\}$ when $i = l$, and then point l is the contaminant source.

If there're two contaminant sources, we use $P_k^i\,(i,\ k = 1, 2, \cdots, 319)$ generated above to calculate the sum pollution concentration caused by two contaminant sources. For example, we assume points 1, 2 are two contaminant sources, and then we get the estimated pollution concentration at point k:

$$P_k^{1,2} = P_k^1 + P_k^2.$$

Similarly, we get

$$\mathrm{MSE}_{i,j} = \frac{\sum_{k=1}^{319} (P_k^{i,j} - P_k)^2}{319},$$

and find out $\min_{i,j}\{\mathrm{MSE}_{i,j}\}$ when $i = m,\ j = n$, and points m and n are the most possible two contaminant sources.

By using the same method we could get the most possible k contaminant sources ($k = 1, 2, 3, \cdots$). But we cannot increase k to 319, which will be too large for programming. So we just increase k until the minimum MSE of $k-1$ assumed sources is smaller than the minimum MSE of k assumed sources. Thus, we get the most possible number of the contaminant sources and the location of them.

By combining the two sub-models, we can predict the location of the contaminant sources.

4 Results & interpreting the results

4.1 Results

According to 'Evaluation Soil Contamination' published by U.S. department of interior[3], it is reasonable to assume that $\tau_{\mathrm{As}} = 0.38$, $\tau_{\mathrm{Cd}} = 0.18$, $\tau_{\mathrm{Cr}} = 0.64$, $\tau_{\mathrm{Cu}} = 0.42$, $\tau_{\mathrm{Hg}} = 0.21$, $\tau_{\mathrm{Ni}} = 0.17$, $\tau_{\mathrm{Pb}} = 0.08$, $\tau_{\mathrm{Zn}} = 0.35$; and we assume that $\sigma = 500$, $\delta = 5000$. The results of the estimated contaminant sources and the related minimum MSEs are listed in Tables 1, 2, 3.

Table 1: Four possible contaminant sources

Four contaminant sources				Minimum MSE
(3299, 6018)	(2383, 3617)	(13694, 2357)	(15248, 9106)	1892.324
(2383, 3617)	(15248, 9106)	(13694, 2357)	(13797, 9621)	2498.879
(15248, 9106)	(13694, 2357)	(9319, 6799)	(2383, 3617)	2754.345
(13694, 2357)	(13797, 9621)	(2383, 3617)	(3299, 6018)	3564.231

Table 2: Five possible contaminant sources

Five contaminant sources					Minimum MSE
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(3299, 6018)	1094.358
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(13797, 9621)	1732.398
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(9319, 6799)	1921.301
(2383, 3617)	(15248, 9106)	(13694, 2357)	(3299, 6018)	(13797, 9621)	2487.628

Table 3: Six possible contaminant sources

Six contaminant sources	Minimum MSE
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (3299, 6018) (13797, 9621)	1108.231
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (3299, 6018) (9319, 6799)	1324.387
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (3299, 6018) (8629, 12086)	1431.216
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (13797, 9621) (9319, 6799)	1442.418

4.2 Interpreting the result

From the mean square error, it is seen that when there are five contaminant sources, we achieve the minimum mean square errors and their locations are (2383, 3617), (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018). This is reasonable because all of these five points are frequently shown in the above tables, indicating that they are of a large probability to be the contaminant points.

However, since the mean square error is still very large, we have to suspect our result. Because we have made an assumption which states that the only possible contaminant points are those points shown in the data file, but in reality, it may not be correct. If we revise it a bit and try the middle points between (2383, 3617) and (2708, 2295) and also substitute the middle point of their heavy metals' concentrations, we can obtain a better mean square error which is 190.134, 5 times smaller than the previous mean square error, when we try 3 contaminant points. These three contaminant points are (15248, 9106), (13694, 2357), (2545.5, 2956).

5 Additional information & evolution model

If time factor can be taken into consideration, we could complete our work to develop an evolution model for the geological environment of the city. Hence we need to collect the concentrations of heavy metals at different times at the same points as we choose previously. Analyzing the dynamic data, we may obtain the accurate σ. As time goes by, we are also able to collect the weather data weather of the city, so as to analyze the impact on pollution distribution by rainfalls, the wind direction and the information about the rivers across the city.

Given the information mentioned above, we could modify our model by correcting the pollution extent and wind direction in diffusion by air sub-model, adding the factor of the rivers and rainfalls. Furthermore, we can develop a new model to study the evolution models of geological environment of the city on the basis of the current model. In the new model, we could analyze the pollution distribution of a certain point as time goes on, instead of the pollution distribution in a large area at a certain time. This new model will also help us predict the trend of the pollution distribution as time goes on. What's more, we could predict what the environment we reside will be in the future, and the significance of the environmental protection will become a highlight of our life. Therefore, we are able to avoid the trend of the environment deterioration, and make our environment better and better.

If there's only one contaminant source, we shall find out $\min_i\{\text{MSE}_i\}$ when $i = l$, and then point l is the contaminant source.

If there're two contaminant sources, we use P_k^i ($i,\ k = 1, 2, \cdots, 319$) generated above to calculate the sum pollution concentration caused by two contaminant sources. For example, we assume points 1, 2 are two contaminant sources, and then we get the estimated pollution concentration at point k:

$$P_k^{1,2} = P_k^1 + P_k^2.$$

Similarly, we get

$$\text{MSE}_{i,j} = \frac{\sum_{k=1}^{319}(P_k^{i,j} - P_k)^2}{319},$$

and find out $\min_{i,j}\{\text{MSE}_{i,j}\}$ when $i = m,\ j = n$, and points m and n are the most possible two contaminant sources.

By using the same method we could get the most possible k contaminant sources ($k = 1, 2, 3, \cdots$). But we cannot increase k to 319, which will be too large for programming. So we just increase k until the minimum MSE of $k-1$ assumed sources is smaller than the minimum MSE of k assumed sources. Thus, we get the most possible number of the contaminant sources and the location of them.

By combining the two sub-models, we can predict the location of the contaminant sources.

4 Results & interpreting the results

4.1 Results

According to 'Evaluation Soil Contamination' published by U.S. department of interior[3], it is reasonable to assume *that* $\tau_{\text{As}} = 0.38,\ \tau_{\text{Cd}} = 0.18,\ \tau_{\text{Cr}} = 0.64,\ \tau_{\text{Cu}} = 0.42,\ \tau_{\text{Hg}} = 0.21,\ \tau_{\text{Ni}} = 0.17,\ \tau_{\text{Pb}} = 0.08,\ \tau_{\text{Zn}} = 0.35$; and we assume that $\sigma = 500,\ \delta = 5000$. The *results of the* estimated contaminant sources and the related minimum MSEs are listed in Tables 1, 2, 3.

Table 1: Four possible contaminant sources

Four contaminant sources				Minimum MSE
(3299, 6018)	(2383, 3617)	(13694, 2357)	(15248, 9106)	1892.324
(2383, 3617)	(15248, 9106)	(13694, 2357)	(13797, 9621)	2498.879
(15248, 9106)	(13694, 2357)	(9319, 6799)	(2383, 3617)	2754.345
(13694, 2357)	(13797, 9621)	(2383, 3617)	(3299, 6018)	3564.231

Table 2: Five possible contaminant sources

Five contaminant sources					Minimum MSE
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(3299, 6018)	1094.358
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(13797, 9621)	1732.398
(2383, 3617)	(15248, 9106)	(13694, 2357)	(2708, 2295)	(9319, 6799)	1921.301
(2383, 3617)	(15248, 9106)	(13694, 2357)	(3299, 6018)	(13797, 9621)	2487.628

Table 3: Six possible contaminant sources

Six contaminant sources	Minimum MSE
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (3299, 6018) (13797, 9621)	1108.231
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (3299, 6018) (9319, 6799)	1324.387
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (3299, 6018) (8629, 12086)	1431.216
(2383, 3617) (15248, 9106) (13694, 2357) (2708, 2295) (13797, 9621) (9319, 6799)	1442.418

4.2 Interpreting the result

From the mean square error, it is seen that when there are five contaminant sources, we achieve the minimum mean square errors and their locations are (2383, 3617), (15248, 9106), (13694, 2357), (2708, 2295), (3299, 6018). This is reasonable because all of these five points are frequently shown in the above tables, indicating that they are of a large probability to be the contaminant points.

However, since the mean square error is still very large, we have to suspect our result. Because we have made an assumption which states that the only possible contaminant points are those points shown in the data file, but in reality, it may not be correct. If we revise it a bit and try the middle points between (2383, 3617) and (2708, 2295) and also substitute the middle point of their heavy metals' concentrations, we can obtain a better mean square error which is 190.134, 5 times smaller than the previous mean square error, when we try 3 contaminant points. These three contaminant points are (15248, 9106), (13694, 2357), (2545.5, 2956).

5 Additional information & evolution model

If time factor can be taken into consideration, we could complete our work to develop an evolution model for the geological environment of the city. Hence we need to collect the concentrations of heavy metals at different times at the same points as we choose previously. Analyzing the dynamic data, we may obtain the accurate σ. As time goes by, we are also able to collect the weather data weather of the city, so as to analyze the impact on pollution distribution by rainfalls, the wind direction and the information about the rivers across the city.

Given the information mentioned above, we could modify our model by correcting the pollution extent and wind direction in diffusion by air sub-model, adding the factor of the rivers and rainfalls. Furthermore, we can develop a new model to study the evolution models of geological environment of the city on the basis of the current model. In the new model, we could analyze the pollution distribution of a certain point as time goes on, instead of the pollution distribution in a large area at a certain time. This new model will also help us predict the trend of the pollution distribution as time goes on. What's more, we could predict what the environment we reside will be in the future, and the significance of the environmental protection will become a highlight of our life. Therefore, we are able to avoid the trend of the environment deterioration, and make our environment better and better.

References:

[1] Beyer W N. Evaluation soil contamination[S]. Department of Interior, Fish and Worldwide Service, Washingtan, 1990

[2] Jiang Y, et al. Contamination, source identification, and risk assessment of polycyclic aromatic hydrocarbons in agricultural soil of Shanghai, China[C]// Environmental Monitoring and Assessment, 2011, 183: 139-150

[3] Lian F, Wang Y X. Single-factor and Nemerow multi-factor index to assess heavy metals contamination in soils on railway side of Harbin-Suifenhe railway in Northeastern China[J]. Applied Mechanics and Materials, 2011, 71-78: 3033-3036

重金属污染源估测模型

李泽喆， 许经纬， 徐文继

(新加坡国立大学数学系，新加坡 119076)

摘 要: 本文提供了一个根据污染数据预测污染源位置的数学模型．首先假定某个特定的污染数据采集点作为污染源，将污染物传播方式分为土壤传播和空气传播；然后运用多种数学工具，综合两种传播方式对污染分布进行模拟；最后通过改变对污染源位置的假设，并将每种假设下的模拟污染分布与污染的真实分布进行比较，从而得到最可能的污染源数量和位置．具体过程由计算机实现．

关键词: 污染源；位置；重金属污染

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0091-07

城市表层土壤重金属污染分析

郑继勇， 翟冬菊， 金木红

指导教师： 雷 宏

(铜仁学院数学与计算机科学系，贵州 铜仁 554300)

编者按：本文建立了一个清晰简练的对流模型来描述重金属污染的传播特征，切中问题的关键，并利用空间坐标的特征曲线、反向计算污染源的位置，方法得当可行．另外，本文相对污染率定义采用了背景值上限体现了作者自己的想法，全文表达简练流畅．本文的不足之处在于第二问做得较薄弱，在分析污染原因时未考虑各元素之间的关联．

摘 要：本文针对如何有效分析城市表层土壤重金属污染特征的问题，分别从各元素浓度的空间分布，重金属污染的主要原因以及污染物的传播特征等方面展开了深入研究．首先，基于所给样本数据，利用克里格插值法得到各种元素的浓度分布，并用软件给出了各种元素的空间分布图．另外，以浓度背景值为参考标准，基于实际浓度的平均值，利用相对污染率，得出各功能区中各种元素的超标程度；利用地累积指数法对各功能区进行环境质量评价，给出了各功能区每种重金属元素的污染等级；进一步，基于各元素的浓度分布图，分析了各种元素的分布差异．结合各环境的差异性，总结出了人类活动是导致各重金属元素污染的主要原因．其次，通过对重金属污染物传播特性的分析，建立起污染物传播的对流模型．利用空间坐标点的特征曲线，反向计算污染源所有可能的位置．进而，通过对所建模型以及数值结果的合理性分析，提出了增加样本容量、改进插值算法、增加时间序列样本数据、建立时间序列演变模型等改进意见，为更好的研究重金属污染物的演变特性提供了新的思路．

关键词：重金属污染；地累积指数法；克里格插值；对流模型

分类号：AMS(2000) 65K05 **中图分类号**：O29 **文献标识码**：A

1 问题重述 (略)

2 符号说明

$i\,(i=1,2,\cdots,5)$:	依次表示生活区、工业区、山区、交通区、公园绿地区；
$j\,(j=1,2,\cdots,8)$:	依次表示元素 As, Cd, Cr, Cu, Hg, Ni, Pb, Zn；
X_{ij}:	第 i 区第 j 种元素的采样平均值；
$I_{\rm geo}$:	地累积指数；
C_n:	元素 n 的实测浓度值；
B_n:	元素 n 的背景浓度值；
$B\max_i$:	第 i 种元素背景浓度的最大值；
P_{ij}:	第 i 区第 j 种元素相对背景浓度最大值的相对污染率．

3 问题分析

首先，针对各重金属元素的分布问题，基于样本浓度数据，可以采用克里格插值算法得出每种重金属元素的浓度分布，利用软件可绘出各种重金属元素在该城区较为直观的空间分布

图[1]. 进而，以背景值为参考标准，求出各种重金属在该区的相对污染率[2]，并结合地累积指数法，可以获得各功能区中每种元素的超标程度及污染等级. 进一步，根据各元素的浓度分布图，通过分析各功能区每种元素的浓度差异，结合各功能区环境的特殊性，得出导致重金属污染的主要原因. 最后，基于重金属污染物的传播特性，可以建立对应的微分对流模型来刻画其传播规律. 进而，利用空间特征曲线的反向演算，得出所有可能的污染源位置.

4 模型建立与求解

4.1 问题一

基于所给的样本数据，利用克里格插值法得到各种元素的浓度分布，并采用软件绘出8种主要重金属元素在整个城区的空间分布图. 鉴于篇幅，我们仅给出As元素的分布图，如图1.

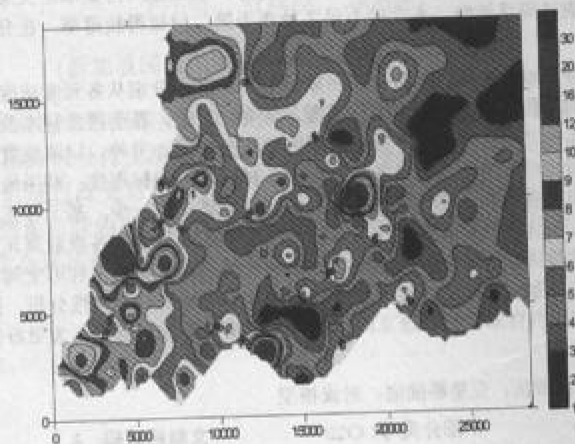


图1: As元素的分布图

通过给定的样本数据，计算各个功能区每种元素浓度的实际均值 X_{ij}，利用公式

$$P_{ij}=\frac{X_{ij}-B_{\max_i}}{B_{\max_i}},\quad j=1,2,\cdots,8,\quad i=1,2,3,4,5, \tag{1}$$

计算出5个区8种元素的相对污染率. 其次，利用地累积指数法公式[3]

$$I_{\mathrm{geo}}=\log_2\frac{C_n}{1.5B_n}, \tag{2}$$

求出5个区8种元素的地累积指数值. 各功能区的计算结果如表1所示. 其中，表中标有'—'符号的数据，即 $P_{ij}\le 0$ 表示该种元素所在功能区没有超标. 表中标有'口'符号的数据，即 $I_{\mathrm{geo}}\ge 1$ 表示该种元素所在功能区已经严重超标，导致了重金属污染程度严重. 从表中数据可以看出，利用相对污染率计算出来的结果和利用地累积指数计算的结果一致性较好. 另外，为了给出各功能区各种元素的综合污染程度，在此对地累积指数进行分级，地累积指数与污染等级的相互关系见表2. 通过对比分析，得出各功能区的污染等级，见表3.

进而可得到各功能区不同元素污染程度大小关系如下：

生活区：Cu > Zn > Hg ≥ Pb ≥ Cr ≥ Cd > As;

工业区：Hg > Cu > Zn > Cd > Pb > As > Cr > Ni；

交通区：Hg > Cu > Zn > Cd > Pb > Cr；

公园绿地区：Hg > Cu > Zn > Cd > Pb > As；

山区：无重金属元素污染.

综合分析，各功能区的污染程度从大到小依次为：工业区、交通区、生活区、公园绿地区、山区.

表1: 各区域各重金属元素的浓度指数

区 域	元素	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
生活区	X_{ij}	6.27	289.96	69.02	49.40	93.04	18.34	69.11	237.01
	P_{ij}	0.16	0.53	0.41	1.42	0.82	-0.08	0.61	1.44
	I_{geo}	0.22	0.57	0.57	1.32	0.83	-0.01	0.57	1.20
工业区	X_{ij}	7.25	393.11	53.41	127.54	642.36	19.81	93.04	277.93
	P_{ij}	0.34	1.07	0.09	5.25	11.60	-0.00	1.16	1.87
	I_{geo}	0.43	1.01	0.20	2.69	3.61	0.10	1.00	1.43
山 区	X_{ij}	4.04	152.32	38.96	17.32	40.96	15.45	36.56	73.29
	P_{ij}	-0.25	-0.20	-0.20	-0.15	-0.20	-0.22	-0.15	-0.24
	I_{geo}	-0.42	-0.36	-0.26	-0.19	-0.36	-0.26	-0.35	-0.50
交通区	X_{ij}	5.71	360.01	58.05	62.21	446.82	17.62	63.53	242.85
	P_{ij}	0.06	0.89	0.18	2.05	7.760	-0.11	0.48	1.50
	I_{geo}	0.08	0.88	0.32	1.65	3.09	-0.07	0.45	1.23
公园绿地区	X_{ij}	6.26	280.54	43.64	30.19	114.99	15.29	60.71	154.24
	P_{ij}	0.16	0.48	-0.11	0.48	1.25	-0.23	0.41	0.59
	I_{geo}	0.21	0.52	-0.09	0.61	1.13	-0.27	0.38	0.58

表2: 污染程度与等级

地累积指数I_{geo}	分级	污染程度	地累积指数I_{geo}	分级	污染程度
$I_{geo} \geq 5$	6	极重污染	$1 \leq I_{geo} < 2$	2	中污染
$4 \leq I_{geo} < 5$	5	重污染–极重污染	$0 \leq I_{geo} < 1$	1	无污染–中污染
$3 \leq I_{geo} < 4$	4	重污染	$I_{geo} < 0$	0	无污染
$2 \leq I_{geo} < 3$	3	中污染–重污染			

表3: 各功能区各元素的污染等级

功能区	各元素的污染情况
生活区	Cu和Zn属于中污染，其余除Ni为无污染外均为无污染–中污染
工业区	Hg属于重污染，Cu属于中污染，Zn属于重污染，Cd, Pb, Zn属于中污染，而As和Cr属于无污染–中污染
交通区	Hg属于重污染，Cu, Zn属于中污染，Cd, Cr, Pb属于无污染–中污染，As, Ni属于无污染
公园绿地区	Hg为中污染，As, Cd, Cu, Pb, Zn为无污染–中污染，Cr, Ni为无污染
山区	无污染

4.2 问题二

从问题一的分析中可知，除山区以外的其它四个功能区中，以Hg, Cu, Zn的污染最为突出；其次，也可以初步确定不同功能区重金属的污染的主要原因[4]，现通过给出的319组数据计算出整个城区8种元素浓度的平均值，如表4.

表4: 整个城区各种重金属元素的浓度均值

元素	As	Cd	Cr	Cu	Hg	Ni	Pb	Zn
总体均值	5.68	302.40	53.51	55.02	299.71	17.26	61.74	201.20
背景均值	3.6	130	31	13.2	35	12.3	31	69
背景值范围	1.8–5.4	70–190	13–49	6.0–20.4	19–51	4.7–19.9	19–43	41–97

从上表可以看出，该城区中Hg元素的背景值范围是19–51 ng/g，平均值是50 ng/g，而样本均值是299.71 ng/g，说明样本均值与背景值相差较大，而且通过前面的分析知道的污染主要在工业区、交通区以及生活区．由此可知，该城区含量值的升高主要是由于工业活动和居民区煤的燃烧的影响．而Cu, Zn, Cd, Pb也主要是工业区从事各种生产活动产生的污水、生活区的各种垃圾和生活污水而导致该元素含量的升高．Ni的含量在背景值的附近，所以该城区基本不受Ni的污染．所以，通过上面的具体分析，并结合各功能区环境的具体特性，得出导致重金属污染的主要因素有：

1) 随着大气沉降进入土壤，而大气中的重金属主要来源于能源、运输、冶金和建筑材料生产的气体和粉尘；

2) 随着污水进入土壤的重金属，而污水的来源更为广泛，例如城市生活污水、石油化工污水、工业矿山污水和城市混合污水等等；

3) 随固体废弃物进入土壤的重金属；

4) 随农用物资进入土壤的重金属．

4.3 问题三

基于对重金属污染物传播特征的分析，建立起重金属污染物的传播模型–对流模型，进而通过反演计算，确定污染源的可能位置．

4.3.1 重金属污染物的传播特征

由问题一所获得的各种重金属元素的浓度空间分布图以及其污染程度，可以看出重金属污染物的传播特征大致符合自然界规律：污染由高浓度向低浓度传播．但由于其重金属元素的特

殊性，其传播速度较慢，且不同重金属的传播速度不同．由此就可以得到污染源的特征：污染元素的含量与周围含量相比，相对较大，即其为区域极值．

4.3.2　建立重金属污染物传播的数学模型

$$\frac{\partial u}{\partial t}+\nabla q=0, \tag{3}$$

其中 u 表示重金属的浓度；∇q 表示传播强度的损失率，由

$$q=-k\cdot u\cdot\nabla h, \tag{4}$$

可得

$$\frac{\partial u}{\partial t}=-k\nabla h\nabla u-ku\nabla^2 h=0. \tag{5}$$

因为重金属污染物的传播不仅与时间 t 有关，而且与空间向量 x 有关．令

$$\frac{dx}{dt}=-k\nabla h, \tag{6}$$

进一步得

$$\frac{du}{dt}=ku\nabla^2 h. \tag{7}$$

进而可解得上述对流模型的解为

$$u(x,t)=U_T\exp\Big(-\int_t^T k\nabla^2 h\big(x(\tau)\big)\mathrm{d}\tau\Big), \tag{8}$$

式 (8) 即为重金属污染物传播的数学规律．

4.3.3　确定污染物的位置

根据所有数据的特征，建立如上微分传递模型来寻找其污染源．由上可知，将其寻找污染源的问题转化为寻找图像空间极值的问题．

基于题设给定的样本值，任意选取一个插值节点 $u(p,T)$，沿 p 的特征曲线反向计算 $u(x(t),t;p)\,(0\le t\le T)$，设定一个污染浓度阀值 u^*，对于任意的 $\bar{t}\,(0\le\bar{t}\le T)$，如果满足如下条件

$$B(\bar{t})=\big\{x(\bar{t})\,|\,u\big(x(\bar{t}),\bar{t};p\big)\ge u^*\big\}, \tag{9}$$

则表明 $\bar{t}$ 就是可能的污染源位置．

数值实验：对应每一种重金属元素，任取一个样本点，通过上述原理，得出各种重金属污染源可能的位置，见表 5．

表 5:　各种重金属污染源的可能位置

元素	样本点 $u(p,T)$	阀值 u^*	污染源可能的位置
As	(15918.88, 10954.90)	15	(18234.36, 10954.09) 交通区
Cd	(11577.37, 8071.43)	500	(20549.83, 11530.62) 生活区，(2604.90, 3747.46) 工业区
Cr	(7235.85, 8071.43)	100	(2894.34, 6630.10) 生活区，(4052.08, 4035.71) 工业区
Cu	(10130.20, 8071.44)	150	(19970.96, 9224.50) 生活区，(4052.08, 3170.92) 工业区
Hg	(6656.98, 7206.64)	500	(1157.74, 3170.92) 交通区，(13892.85, 9512.77) 工业区
Ni	(6078.12, 6918.37)	20	无
Pb	(11287.93, 3747.45)	150	(14182.28, 7994.91) 生活区，(4052.08, 3170.92) 工业区
Zn	(8393.59, 6341.84)	400	(9840.76, 3747.45) 交通区，(13024.54, 10605.82) 生活区

基于样本点的数据，通过(8)式进行反演计算，得出各种重金属污染源的可能位置．由表5可以看出，各种重金属污染源主要集中在工业区和生活区，且污染源处的重金属浓度普遍较高．但是也有例外，比如Pb元素，得到其中一个可能的污染源位置为(14182.28, 7994.91)，属于生活区，而该点处的浓度值为38.78，低于该种元素的平均浓度浓度61.7．出现这种现象可能的原因是该种元素所在区域产生了富集现象．另外，Ni元素并没用找到污染源，这可能和整个城区Ni元素本身并没超标有关．

4.4 问题四

4.4.1 模型评价

本文基于所给样本数据，得到重金属的空间分布情况，并结合重金属污染物的传播特性，建立起重金属污染物的微分传播对流模型，进而根据所给的样本点，通过反演计算，确定了可能的污染源位置，为有关重金属污染问题的解决提供了可行的决策方案．但是，由于所给样本信息有限，以及模型建立过程中的假设理想化，问题分析还存在诸多不足之处．

1) 本文基于克里格插值法，得到了各种重金属元素浓度的空间分布图．然而，由于采样数据过于离散化，数据量小等问题，导致所得空间分布可能存在很大误差，结果的准确性受到很大影响．

2) 本文基于对题设所给数据的分析，分别对各重金属元素污染的主要原因进行了定量、定性分析。但是，分析过于简单，未考虑各元素之间的关联性．

3) 本文基于重金属污染物的传播特性，通过简化假设，建立起其传播的数学模型．由于污染物传播本身的复杂性，导致污染源的位置确定存在很大误差．

4.4.2 建议进一步收集的信息及改进思路

1) 增加样本数据容量．通过增加样本数据容量，从而提高插值精度，为准确判断各空间位置的金属浓度提供可靠的依据．

2) 增加其它一般元素的浓度．进行比较分析研究[5]．

3) 增加时间序列上的样本数据．例如，测量连续五年同一空间坐标处的重金属浓度值．可考虑建立时间序列模型．通过此模型不仅可以很好的确定污染源的位置，还可以很好的研究重金属的传播以及城市地质环境的历史演变规律．

4) 增加受污染的环境变化、病历等数据．为更全面的分析提供准确可靠的依据．

参考文献：

[1] Addison. Surfer lesson[OL]. http://wenku.baidu.com/view/8623e12a4b73f242336c5f64.html

[2] 姜启源，等. 数学模型(第3版)[M]. 北京：高等教育出版社，2003
Jiang Q Y, *et al.* Mathematical Model (3rd Edition)[M]. Beijing: Higher Education Press, 2003

[3] 彭景，等. 地累积指数法及生态危害指数法在土壤重金属污染种的应用及探讨[J]. 广东微量元素科学，2007, 14(8): 13-17
Peng J, *et al.* Application of the index of geo-accumulation index and ecological risk index to assess heavy metal pollution in soils[J]. Guangdong Trace Elements Science, 2007, 14(8): 13-17

[4] 姚德，等. 青岛城区土壤重金属环境地球化学研究[J]. 中国地质，2008, 35(3): 539-549
Yao D, *et al.* Environmental geochemistry of heavy metals in urban soils of Qingdao city[J]. Geology in China, 2008, 35(3): 539-549

[5] 王玉梅. 土壤重金属污染与防治[J]. 安徽农学通报，2005, 11(7): 46-47
Wang Y M. Soil heavy metal pollution and its prevention and cure[J]. Anhui Agricultural Science Bulletin, 2005, 11(7): 46-47

Research on Heavy Metal Pollution in Urban Topsoil

ZHENG Ji-yong, ZHAI Dong-ju, JIN Mu-hong
Advisor: LEI Hong

(Department of Mathematics and Computer Science, Tongren University, Tongren, Guizhou 554300)

Abstract: Aiming at solving the problem to analyze the intrinsic characteristics of heavy metal pollution in the urban topsoil, this paper studies the spatial distribution of the element concentration, the main causes underlying heavy metal pollution and the propagating character of the pollutions. First, based on the sample data, we obtain the concentration distribution of each element by using Kriging interpolation method. And the spatial distribution map is further drawn by virtue of the Surfer software. With reference to the background values of concentration, the degree exceeding the stipulated standards of the various elements is attained in all functional zone based on the average actual concentration by relative contamination rate. In addition, we assess the environmental quality with respect to the geo-accumulation index and provide the pollution levels of each heavy metal element in different functional zone. Moreover, the distribution difference of various elements is analyzed based on the concentration distribution map. Based on such environmental difference, it is evident that human activities are the main cause of heavy metal pollution. Second, we establish the convection model for describing the pollutant spreading by virtue of the propagating character of heavy metal pollutant. All possible locations of polluting sources are calculated reversely using the characteristic curves of space coordinates. After analyzing the intrinsic principal underlying the model and evaluating the experimental results, we put forward some further research orientations, such as the increase of the sample size, the improvement of the interpolation algorithm, and the establishment of the time series evolution model provide some new ideas.

Keywords: heavy metal pollution; the index of Geo-accumulation; Kriging interpolation; convection model

第28卷 增刊1　　工 程 数 学 学 报　　Vol. 28 Supp.1
2011年11月　　CHINESE JOURNAL OF ENGINEERING MATHEMATICS　　Nov. 2011

文章编号:1005-3085(2011)07-0098-07

交巡警服务平台设置与调度方案研究

叶云佳， 刘 剑， 王 禹
指导教师： 曹华林
(海军航空工程学院青岛校区航空机械系，青岛 266041)

编者按：本文从图论角度分析了交巡警平台的设置问题，建立的优化模型表述较为清楚，思路清晰，抓住了问题的实质，对每一问都给出了合理的结果，易于理解，但方法较单一.

摘　要：本文通过研究和分析某城市的交巡警服务平台的分布现状，为了提高交巡警资源的利用率、加快交巡警服务平台的反应速度和平均分配各平台之间的工作强度，利用Floyd算法，最小二乘法和多目标规划，对交巡警服务平台的管辖范围进行了合理的规划. 然后，进一步分析了该市现有警力分配状况中存在的不合理情况，并结合实际情况提出了改进方案，重新分配了警力，并根据具体情况，制定了封锁路口的方案.

关键词：Floyd算法；最小二乘法；多目标规划

分类号：AMS(2000) 90C11　　中图分类号：O221　　文献标识码：A

1　问题重述 (略)

2　模型假设

1) 该城市的所有道路均为双行道，每条道路均为直线；
2) 不考虑道路交通状况对出警速度的影响；
3) 一个交巡警服务平台管辖区域内的发案率 (次数) 总和与该平台工作强度成正比；
4) 假设犯罪嫌疑人驾车逃跑的速度与警车出警的车速相同，都为60 km/h.

3　符号说明

d_{ij}：第i路口到第j路口之间的距离；

n_j：第j路口的发案率 (次数)；

g_i：第i平台处理案件数的总和 (即工作量)；

$\bar{g}$：某区各平台平均处理的案件数 (即平均工作量).

4　模型的建立与求解

4.1　模型一：交巡警服务平台管辖范围确定问题

4.1.1　模型准备

1) 各路口之间距离的确定

假设A区的交通网络与平台设置的示意图中A路口的坐标为(x_1, y_1)，B路口的坐标为(x_2, y_2)且A、B之间有道路连通，根据给出的比例尺，两路口距离即为

$$d_{AB} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \times 0.1\,(\mathrm{km}). \tag{1}$$

2) 利用Floyd算法[1]求解出每两个路口之间的最小路程

假设图G权的邻接矩阵为A_0，经过计算，我们发现各交巡警服务平台都无法在3min内到达的路口只有28, 29, 38, 39, 61, 92这六个路口.

4.1.2 模型约束

1) 0-1变量约束．f_{ij}为表示第i路口的平台是否管辖第j路口的0-1变量，即

$$f_{ij} \in \{0, 1\}. \tag{2}$$

2) 每个路口都必须被管辖的约束．要求每一个路口由且仅由一个平台管辖，即

$$\sum_{i=1}^{20} f_{ij} = 1. \tag{3}$$

3) 设置了平台的路口的约束．对于有平台的路口，应直接由该平台进行管辖，不应考虑其他情况，即

$$f_{ij} = 1. \tag{4}$$

4) 尽量3min内赶到事发地的约束．除了6个特殊的点以外，其他的点都要满足3min内有交巡警赶到，即到交巡警服务平台的距离应小于等于3km，因此有

$$f_{ij} \times d_{ij} \le 3. \tag{5}$$

5) 对于6个3min内不能到达的路口，应该由与其距离最近的路口的平台来管理，即

$$d_{ij} \times f_{ij} \le \min_i (d_{ij}). \tag{6}$$

4.1.3 模型建立

我们建立各交巡警服务平台工作量方差最小二乘法规划模型

$$\min \quad \frac{1}{20} \sum_{i=1}^{20} (g_i - \bar{g})^2 \tag{7}$$

$$\text{s.t.} \begin{cases} g_i = \sum\limits_{j=1}^{92} (f_{ij} \times n_j), \\ \bar{g} = \dfrac{\sum\limits_{j=1}^{92} n_j}{20}, \\ \sum\limits_{i=1}^{20} f_{ij} = 1, \\ f_{ii} = 1, \\ f_{ij} \times d_{ij} \le 3, \quad j \ne 28, 29, 38, 39, 61, 92, \\ f_{ij} \times d_{ij} \le \min\limits_i (d_{ij}), \quad j = 28, 29, 38, 39, 61, 92, \\ f_{ij} \in \{0, 1\}, \\ i = 1, 2, \cdots, 20, \quad j = 1, 2, \cdots, 92. \end{cases} \tag{8}$$

4.1.4 模型求解

我们先通过Matlab编程利用Floyd算法[1]，求解得到任意两点之间的最短路程矩阵；然后利用Lingo优化编程得到各交巡警服务平台的管辖范围．计算得到这种划分方案各个交巡警服务平台管辖地区发案率的方差为3.02.

4.2 模型二：交巡警服务平台警力封锁调度问题

1) 模型二建立

我们考虑到封锁路口时，不仅要最后到位的平台时间尽可能短，而且要求其他平台的速度也要尽可能的快，即各平台到位的时间总和最小．因此，在第一层模型的基础上，第二层规划模型将目标函数改为各个出警封锁的平台所用的时间总和最小．

第一层目标函数：

$$\min \quad \max\left(f_{ij}\times\frac{d_{ij}}{v}\right). \tag{9}$$

第二层目标函数：

$$\min \quad \sum_{i=1}^{20}\sum_{j} f_{ij}\times\frac{d_{ij}}{v}. \tag{10}$$

约束条件：

$$\text{s.t.}\begin{cases}\sum\limits_{i=1}^{20} f_{ij}=1,\\ \sum\limits_{j} f_{ij}\le 1,\\ f_{ij}\in\{0,1\},\\ i=1,2,\cdots,20, \quad j=12,14,16,21,22,23,24,28,29,30,38,48,62.\end{cases} \tag{11}$$

2) 模型二的求解

通过计算，按照上述方案能够在8.02min内，以最快速度完成对所有进出A区的交通要道路口的封锁．

4.3 模型三：新增交巡警平台布置问题

1) 解题思路及建模准备

经过上述计算我们得到交巡警3min内不能够到达(即大于3km)的路口有28, 29, 38, 39, 61, 92号6个路口．其中能够满足到28, 29号路口的路程小于等于3km的路口为28号和29号路口；满足到38, 39号路口的路程小于等于3km的路口有38, 39, 40号路口；满足到达61号路口好路口的路程小于等于3km的路口有48, 61号路口；满足到达92号路口的路程小于等于3km的路口有87, 88, 89, 90, 91, 92号路口．因此，我们把不能满足出警时间大于3min的点归为4组，在每组内只要再新增一个交巡警服务平台，就能够满足出警时间均小于等于3分钟的要求，所以至少需要再新增4个交巡警服务平台．

2) 模型三的建立

这里我们仍然采用多目标规划的方法，在工作量最大平台的工作量尽可能小的前提下，使工作量的分配尽可能得平衡．

第一层目标函数：

$$\min \quad \max\left(\sum_{j=1}^{92} f_{ij}\times n_j\right). \tag{12}$$

第二层目标函数：

$$\min \quad \frac{1}{24}\sum_{i=1}^{24}(g_i-\bar{g})^2. \tag{13}$$

约束条件：

$$\text{s.t.}\begin{cases} f_{ij}\in\{0,1\},\\ d_{ij}\times f_{ij}\le 3,\\ \sum\limits_{i=1}^{24} f_{ij}=1,\\ i=1,2,\cdots,24,\\ j=1,2,\cdots,92. \end{cases} \tag{14}$$

3) 模型三的求解

首先我们用Lingo编程求解出第一层规划模型的结果为

$$\max\left(\sum_{j=1}^{92} f_{ij}\times n_j\right)=8.5.$$

然后将这个结果作为约束，代入第二层规划模型中，再用Lingo编程求解，得到各交巡警服务平台的管辖范围，其中A21平台配置在第28号路口或第29号路口均可，A22平台配置在第40号路口，A23平台配置在第48号路口，A24号平台配置在第87号路口.

这种分配方案各交巡警服务平台管辖地区发案率的方差为2.20，与没有新增平台之前的3.02相比有一定的改进，工作量的分配更加趋于平衡.

4) 模型三的改进

由上面的模型计算结果我们可以看出：新增4个交巡警服务平台的方案虽然对各平台工作量的平衡有一定的促进作用，但平衡作用并不是很明显，最大工作量仍为8.5，因此我们可以考虑新增5个交巡警服务平台的方案.

经过建模、编程、求解，我们决定将第5个新增的平台A25设置在第22号路口，这种分配方案各交巡警服务平台管辖地区发案率的方差为1.51，最大工作量为6.1，明显小于新增4个平台的结果2.20和8.5，说明各交巡警服务平台的平衡性得到显著的提升.

4.4 模型四：全市现有交巡警服务平台设置合理性及改进问题

4.4.1 对全市现有交巡警服务平台设置合理性评价

1) 服务平台警力大部分集中在A区，警力充足，而D区和F区的警力明显不足；

2) C区和F区平均每个平台处理的案件数较多，服务平台的工作强度较大；

3) D区交巡警服务平台管辖的面积大，人口多，而平台数却相对较少；

4) 我们发现除了A区各平台3min内都无法到达的路口较少外，其他各区到每个平台的距离都大于3km的路口很多，B区有6个，C区有43个，D区有12个，E区有30个，F区有28个，这就会造成到某些路口的出警时间过长，无法满足迅速到达案发地点的要求.

4.4.2 全市警力重新配置模型的建立

1) 建模准备

我们采用加权的方法来确定重新分配后每个区的平台数量. 我们确定影响交巡警服务平台分布的因素主要有各区路口总数，各区人口数和各区面积数.

我们采用极值标准化的方法对附表中给出的数据进行标准化，我们将标准化后的数值与对应的权值相乘，得到加权求和后的结果为：A区为0.62634，B区为0.38877，C区为0.842281，D区为0.592818，E区为0.807227，F区为0.69527。

将上述数据换算成百分比为：A区占15.9%，B区占9.8%，C区占21.3%，D区占15%，E区占20.4%，F区占17.6%．由于警务平台总数为80，所以重新分配的警务平台数依次为13, 8, 17, 12, 16, 14．

2) 具体规划模型的建立

对于每个城区交警平台布置位置的确定以及每个平台具体管辖范围的划分，我们是使用0-1规划模型进行最优方案的选取．下面我们以B区交巡警服务平台的重新规划为例进行规划模型的建立．

目标函数：

$$\min \quad \max\Big(\sum_{j=93}^{165} f_{ij}\times\frac{d_{ij}}{v}\Big). \tag{15}$$

约束条件：

$$\text{s.t.}\begin{cases} f_{ij}\in\{0,1\},\\ k_i\in\{0,1\},\\ \sum\limits_{i=93}^{165} k_i=8,\\ \sum\limits_{i=93}^{165} f_{ij}=1,\\ \text{若}\ \sum\limits_{j=93}^{165} f_{ij}\neq 0,\quad \text{则}\ k_i=1,\\ f_{ii}=k_i,\\ i=93,94,\cdots,165,\\ j=93,94,\cdots,165. \end{cases} \tag{16}$$

3) 模型四的求解

我们利用Lingo编程求解，求解出交巡警服务平台在各区分布，如表1所示．

4.5 模型五的建立

1) 模型五的准备

由于我们知道犯罪嫌疑人逃跑的起点和时间，并假设其车速与警车速度相当，则我们可以确定犯罪嫌疑人车辆在某一时间可能出现的地区范围，实现用最小的包围圈对犯罪嫌疑人的围堵．我们规定满足下列条件，就可认为交巡警能成功围堵住犯罪嫌疑人的车辆：

$$\frac{dj(i)}{v}\leq\frac{df(i)}{v}-\Delta t, \tag{17}$$

其中$dj(i)$表示警务平台到i路口的最短路程；$df(i)$表示犯罪分子到i路口的最短路程，$\Delta t=3\,\text{min}$．

通过对全体数据的分析，我们将不能够围堵住犯罪嫌疑人的路口全部选出来作为最初计划围堵的范围，这些路口为：7, 8, 9, 30, 31, 32, 33, 34, 35, 36, 39, 45, 46, 47, 48, 61．在这里，我们引入集合A, B (A, B集合之间均无重复的元素)：

$$A=\{A_1,A_2,\cdots,A_k\},\quad p\in A,\qquad B=\{B_1,B_2,\cdots,B_k\},\quad q\in B, \tag{18}$$

其中 A_i 为集合 A 的子集，表示在犯罪嫌疑人车辆路过第 i 个路口之后，可能选择未被警方封锁的路口编号组成的集合；B_i 为集合 B 的子集，表示在犯罪嫌疑人车辆路过第 i 个路口之后，可能选择的被警方封锁的路口的编号组成的集合．集合 A, B 最后一个子集的编号均为k．

表1:　警力重新分配后各区平台分布表

全市六城区	分配平台数	重新分配平台的位置(路口编号)
A	13	1, 3, 5, 10, 11, 12, 13, 14, 29, 31, 40, 60, 91
B	8	94, 95, 97, 98, 100, 103, 125, 152
C	17	167, 168, 169, 171, 174, 175, 176, 177, 178, 180, 181, 200, 210, 239, 262, 275, 300
D	12	327, 328, 331, 332, 339, 343, 344, 360, 362, 363, 368, 370
E	16	372, 373, 374, 377, 379, 381, 383, 384, 385, 387, 388, 390, 396, 404, 419, 457
F	14	477, 478, 479, 480, 482, 483, 484, 485, 505, 541, 549, 574, 579, 582

2)　模型五的建立

目标函数：

$$\min \quad \mathrm{account}\,(A). \tag{19}$$

约束条件：

$$\begin{cases} A_k = \emptyset, \\ \sum\limits_{i} f_{iq} = 1, \\ \sum\limits_{q \in B} f_{iq} = 1, \\ \frac{d_{in}}{v} \le \frac{df_{in}}{v} - \Delta t, \\ i = 1, 2, \cdots, 485. \end{cases} \tag{20}$$

3)　模型五的求解

通过求解，我们求得需要派遣22个平台进行围堵，account $(A) = 45$，其中最大的 d_{iq} 为8.79 km，即能够在8.79min内实现对犯罪嫌疑人车辆的围堵，此时犯罪嫌疑人的车辆最多经过了45个路口．

5　模型评价

由于题中信息有限，所以本文模型在实际应用时仍存在以下不足：

1)　在最后一问的求解中，我们假设犯罪嫌疑人逃跑时车速为60km/h，但如果其车速加快，则不能根据上述方法求解；

2)　对犯罪嫌疑人包围的方案只涉及到了包围，而没有考虑如何抓捕，可能会使得抓捕的时间比较长．

参考文献:

[1] 韩中庚. 数学建模方法及其应用[M]. 北京：高等教育出版社，2009
Han Z G. Mathematical Modeling Method and its Application[M]. Beijing: Higher Education Press, 2009
[2] 宋叶志，贾东永. Matlab数值分析与应用[M]. 北京：机械工业出版社，2009
Song Y Z, Jia D Y. Matlab Numerical Analysis and Application[M]. Beijing: Mechanical Industry Press, 2009

Research on Setting and Scheduling Scheme of Traffic and Patrol Police Service Platform

YE Yun-jia, LIU Jian, WANG Yu
Advisor: CAO Hua-lin

(Aeronautical Department of Mechanical Engineering, Naval Aeronautical Engineering Institute, Qingdao Campus, Qingdao 266041)

Abstract: The paper investigates the distribution situation of the traffic and patrol police platform in a certain city. We develop a reasonable programming for traffic and patrol police service platform that needs to raise the resource utilization rate and enforce its response speed by using the Floyd algorithm in grapy theory as well as the least square and multiple objective programming. Furthermore, we investigate the unreasonable factors underlying the current police scheduling condition, and by combining with the practical situation we further propose the advanced scheme that re-schedules the police strength. Besides, we establish the scheme of blockading crossing roads based on specific situations.

Keywords: Floyd algorithm; least square method; multiple objective programming

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0105-12

交巡警服务平台的设置与调度模型

但 琦[1]， 韩中庚[2]， 杨廷鸿[1]

(1- 解放军后勤工程学院基础部，重庆 401311； 2- 解放军信息工程大学信息工程学院，郑州 450002)

摘 要: 本本文针对2011年全国大学生数学建模竞赛的B题“交巡警服务平台的设置与调度”问题，首先介绍了问题的立意与背景，然后分析给出了相关几个具体问题的解决方法和具体的模型，以及问题的主要求解方法和结果，接着简要分析了参赛论文中普遍存在的主要问题，最后提出了值得进一步研究的几个问题.

关键词: 交巡警服务平台；交通网络图；指派模型；围堵模型

分类号: AMS(2000) 90C30 **中图分类号**: O221.2 **文献标识码**: A

1 问题的背景与立意

“交巡警服务平台的设置与调度”问题作为2011年全国大学生数学建模竞赛本科组的B题[1]，主要是根据重庆市交巡警平台设置的实际情况抽象加工而成的.

早在2010年2月7日，重庆市公安局交通巡逻警察总队挂牌成立，之后部署了300个交巡警平台、8300名警力，肩负起“破案打击、治安防控、交通管理、服务群众”四大职能. 重庆市的交巡警服务平台，看上去就是一个占地约10平方米、高7米的“小亭子”. 每个平台配备有1辆巡逻警车、4辆摩托车、GIS、GPS定位系统、视频监控镜头、移动查询执法终端、360度循环录像、高空定位爆闪灯、施救设施、防暴器材，集众多先进警务技术于一体. 平台通过3G网络及专线与“大情报”平台、警务综合应用平台、社会信息资源平台和视频监控系统无缝对接，可随时查询比对人口、逃犯和车辆信息. 平台还配备冰箱、微波炉、冷暖设备、安全警戒线和防雷设施，民警配备防寒靴、发热背心，兼顾了人性化、安全性和舒适性.

交巡警不仅在形式上实现了110与122的合警联动，更建立了高效的指挥机制. 警方接处警速度从过去的5到10分钟缩短为3到5分钟. 在主城区任何一个地点，能够在3分钟内集结5台警车、20台警用摩托车、60名警力，实现对警情“第一时间、第一现场、第一信息、第一处置”.

交巡警平台不但能方便群众、缩短警民距离，也能预防犯罪、缩小犯罪空间. 它替代了原有的上千个临街占道、老式破旧的治安岗亭和警务室，布设于案件高发、防范薄弱的重点部位，依托“大情报”系统，能紧盯刑事案高发时段、地段和交通管理乱点、堵点、事故多发点，做到警力跟着警情走，勤务伴同民意行. 交巡警制度整合了警力资源，构成了刑事执法、治安管理、交通管理、服务群众四大职能有机融合的新型防控体系. 为了更有效地贯彻实施这些职能，需要在市区的一些交通要道和重要部位设置交巡警服务平台. 由于警力资源是有限的，如何根据城市的实际情况与需求合理地设置交巡警服务平台、分配各平台的管辖范围、调度警力资源，是警务部门面临的一个实际课题.

这个问题我们最早是2010年4月在重庆的交巡警平台首次注意到，当时觉得交巡警平台很神秘，出于好奇提出了相关的一些问题. 例如：我们发现交巡警平台有的相距较远，有的相距较近，那么它们的位置是根据什么来选择确定的？一个平台分管多少地方？各平台之间是如何分工，又如何合作的？发生突发事件后是如何调度的？带着这些问题，当2010年8月再次到重

庆开会时，我们专门到交巡警平台去调研，此后又走访了交巡警总队进行咨询，收集了大量的相关数据资料和重庆市最新的多种版本地图等．直到2010年底讨论形成了一个初步的问题，并给出了问题解决的思路．后来经过全国的命题研究会，根据专家组的意见，对题目又进行进一步的加工，直到2011年5月完成了基本问题的简要解答．此后又与专家组相关专家多次交流修改和完善，直到竞赛最终完整地确定了这个题目．

2 A区交巡警服务平台的相关问题

这部分针对A区范围内的交巡警平台研究解决下面三个问题：

1) 要求合理地为各个交巡警服务平台分配管辖范围，尽量在3分钟内平台交巡警能够到达各管辖的路口；

2) 若需要对进出该区的13条交通要道实施全封锁，则要求给出合理的平台调度方案；

3) 根据现有的资料和数据，确定需要在该区增设平台的个数和具体设置的位置方案．

2.1 A区平台的管辖范围的合理分配问题

2.1.1 A区交通网络赋权图和最短路矩阵

将A区的交通线路抽象为交通网络赋权图．用x_i表示第$i\,(i=1,2,\cdots,m)$个路口，y_j表示第$j\,(j=1,2,\cdots,n)$个交巡警服务平台．以路口为节点，路口之间的公路为边，其公路的长为对应边的权重，于是就可以建立A区的一个交通网络赋权图．将相应的邻接矩阵记为$L=(l_{ij})_{m\times m}$，根据网络优化中求最短路问题的Floyd算法[2,3]，用Matlab编程计算出任意两个节点之间的最短距离，记相应的最短路矩阵为$A_{ij}\,(i,j=1,2,\cdots,m)$，单位：km．

2.1.2 A区平台管辖范围的优化模型

要确定A区各平台的管辖范围，就是将A区内的每一个交通路口合理地分配给一个指定平台管辖的方案．这里所说的合理性主要是体现在两个方面：

1) 尽量在3分钟内平台交巡警能够到达各管辖的路口，即要求每个平台到达所有管辖路口的最大时间尽量小；

2) 从实际出发，要求各平台的出警工作量应尽量均衡．

为此，构造决策矩阵$X=(X_{ij})_{m\times n}$，其中决策变量为

$$X_{ij}=\begin{cases}1, & \text{若路口 } x_i \text{ 由平台 } y_j \text{ 管辖},\\ 0, & \text{其它},\end{cases}$$

用B_{ij}表示A区内路口x_i到平台$y_j\,(i=1,2,\cdots,m,\ j=1,2,\cdots,n)$的最短路程(或时间)，即$B=(B_{ij})_{m\times n}$．在决策矩阵$X$下，该区内各个路口到达其管辖平台的最短路程(或时间)矩阵为

$$T=(T_{ij})_{m\times n}=(X_{ij}\cdot B_{ij})_{m\times n}.$$

于是，对于平台y_j来说，最大出警时间为

$$\max_{1\le i\le m} T_{ij},\quad j=1,2,\cdots,n.$$

另一方面，将A区内各个路口发案量记为向量$W=(w_1,w_2,\cdots,w_m)$，其中w_i表示A区内路口$x_i\,(i=1,2,\cdots,m)$发案量，则各平台的工作量可表示为$G=W\cdot X=(G_1,G_2,\cdots,G_n)$．

为了使各平台的工作量尽量均衡，则要各平台工作量的标准差最小，即要求

$$\sigma(G)=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(G_j-\overline{G}\right)^2}$$

的最小，其中 $\overline{G}$ 为平均工作量.

综上所述，以A区内所有平台的最大出警时间的最小和各个交巡警平台工作量标准差最小为目标函数，建立各平台管辖范围分配的双目标优化模型如下：

$$\min_X \max_{\substack{1\le i\le m\\1\le j\le n}} T_{ij},\quad \min_X \sigma(G),$$

$$\text{s.t.}\begin{cases}\sum\limits_{j=1}^{n}X_{ij}=1, & i=1,2,\cdots,m,\\ \sum\limits_{i=1}^{m}X_{ij}\ge 1, & j=1,2,\cdots,n,\\ X_{ij}=0\text{ 或 }1, \quad \text{且当 } i=j \text{ 时}, & X_{ij}=1,\\ i=1,2,\cdots,m, \quad j=1,2,\cdots,n.\end{cases}$$

其中第一个约束条件为每一个路口有且仅有一个平台对其管辖，第二个约束为每个平台至少管辖一个路口及附近区域，第三个约束为决策变量，且平台所在的路口由该平台管辖.

2.1.3　模型的求解与结果分析

对于上面的双目标0-1规划模型，直接求解是比较困难的．实际上，通过直观分析不难看出，有6个路口在3分钟内是无法达到的，即为28与29, 38, 39, 61和92号路口，为此，按就近原则分别直接分配给15号、16号、2号、7号和20号平台管辖．除此之外的路口都可以满足在3分钟内到达的要求，即可将约束条件

$$\max_{\substack{1\le i\le m\\1\le j\le n}} T_{ij}\le 3$$

加入模型中，则将其转化为以各平台工作量均衡指标(标准差)最小为目标的0-1规划模型.

另一方法，将两个目标作线性加权和化为单目标问题求解，在排除了6个3分钟内不能到达的路口后，最大出警时间就不再是主要的目标了，即对最大出警时间目标赋以较小的权值，而对工作量均衡目标赋以较大的权值．用Lingo软件直接求解，或用Matlab编程求解都可以得到相应的结果．求解方法不同，所取权值不同，其求解结果略有差别.

一种可行分配方案的结果：20个平台平均最大出警时间为2.1759分钟，最大时间为5.7分钟；平均每天的出警次为6.225次，最多的为11.5次，最少为1.6次；不能在3分钟内到达的有6个路口.

2.2　A区20个平台对13个进出路口的全封锁模型

从20个平台中选择13个对所要封锁的目标路口进行一对一的封锁，用 $(d_{ij})_{20\times13}$ 表示A区内平台 $y_i\ (i=1,2,\cdots,20)$ 到目标路口 $z_j\ (j=1,2,\cdots,13)$ 的最短路(时间)矩阵．平台 y_i 对目标路口 z_j 进行封锁的决策矩阵为 $X=(x_{ij})_{20\times13}$，其中当平台 y_i 封锁路口 z_j 时，$x_{ij}=1$，其它 $x_{ij}=0$.

在决策矩阵 X 下，各平台到达要封锁目标路口的最短路程(时间)矩阵为 $T=(T_{ij})_{20\times13}=(d_{ij}x_{ij})_{20\times13}$，完成封锁的最大时间为 $\max\limits_{\substack{1\le i\le 20\\1\le j\le 13}} T_{ij}$，要实现全封锁的目标是最大时间最小，则问

题可归结为一个不完全的指派问题[2]，其优化模型为：

$$\min Y=\max_{\substack{1\le i\le 20\\ 1\le j\le 13}} T_{ij},$$

$$\text{s.t.}\begin{cases}\sum\limits_{j=1}^{13} x_{ij}\le 1, & i=1,2,\cdots,20,\\ \sum\limits_{i=1}^{20} x_{ij}=1, & j=1,2,\cdots,13,\\ x_{ij}=0\ 或 1, & 且\ x_{jj}=1,\\ i=1,2,\cdots,20, & j=1,2,\cdots,13.\end{cases}$$

其中第一组约束表示一个平台最多只能封锁一个目标路口；第二组约束表示每个目标路口有且只能由一个平台封锁．用Lingo软件直接求解可以得到最优的结果，最优的封锁调度方案和相应的封锁时间，如表1所示．

表1: A区的全封锁调度方案(时间单位：分钟)

平台	封锁路口	封锁时间	平台	封锁路口	封锁时间	平台	封锁路口	封锁时间
A2	38	3.9822	A9	16	1.5325	A14	21	3.265
A4	62	0.35	A10	22	7.7079	A15	28	4.7518
A5	48	2.4758	A11	24	3.8053	A16	14	6.7417
A7	29	8.0155	A12	12	0			
A8	30	3.0608	A13	23	0.5			

要实现全封锁的最长时间为8.0155分钟，总平均封锁时间为3.55296分钟．实际上，这个最优的封锁方案结果应该是唯一的．

2.3 A区需要增加平台的设置模型

在实际中，要新增设一个平台必然会需要一定的建设和运行成本，将这些成本视为投入，新增平台后，将需要对全区各平台的管辖范围进行重新分配，从而会使得所有平台在辖区内的最大出警时间和相应工作量指标发生改变，将这种改变的效果视为产出，那么这个问题就可以视为一个"投入与产出"的问题．

在现有平台的基础上，因为没有设置平台的路口节点有 $m-n$ 个，即考虑在这 $m-n$ 个可能的节点中任取一个增设平台，按照问题(1)的方法，对增加平台后各平台的管辖范围进行重新分配，则一定存在一个增设平台的方案，使得"产出"有最大值．此方案就是增设一个平台的最优方案．事实上，只要给出增设一个平台的模型和求解方法，则增设两个、三个、四个和五个的情况可以同理可得．

对没有设平台的路口进行重新排序，并记为 $x_1,x_2,\cdots,x_{m-n}$，则可能要增设一个平台位置的决策变量记为 $(r_1,r_2,\cdots,r_{m-n})$，其中若在路口 x_i 设置新平台 $r_i=1$，其它 $r_i=0$．

考虑增加一个平台，则平台的个数由原来的 n 个变为 $n+1$，由问题(1)的方法，对 $n+1$ 个平台的管辖范围进行重新分配，并计算所有平台的最大出警时间，记为 t_1，相应的工作量标准差为 σ_1．记增设平台之前所有平台的最大出警时间为 t_0 和工作量的标准差为 σ_0．定义增设一个平台后的产出值(增设效益)为

$$F=\rho\frac{t_0-t_1}{t_0}+(1-\rho)\frac{\sigma_0-\sigma_1}{\sigma_0},$$

其中ρ为经验比例系数．于是，可得如下增设平台的优化模型：

$$\max F=\rho\frac{t_0-t_1}{t_0}+(1-\rho)\frac{\sigma_0-\sigma_1}{\sigma_0},$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{m-n} r_i=1,\\ r_i=0\text{或}1,\quad i=1,2,\cdots,m-n.\end{cases}$$

对于模型的求解，参照问题(1)的方法，则可以得到增设一个平台的位置和产生的效益值．在考虑增设第一个平台之后，再考虑增设第二个平台，依此类推．不妨取$\rho=0.5$，讨论增加1至5个平台后所得到效果情况比较结果，如表2所示．

表2: 增设新平台后对应效果(时间单位：分钟)

增设平台情况	出警时间	标准差	产出效益	增设平台情况	出警时间	标准差	产出效益
无增设平台	5.7005	2.6606	—	增设3个平台	3.6822	2.4157	0.045802
增设1个平台	4.1902	2.9545	0.077239	增设4个平台	3.6822	1.9962	0.086828
增设2个平台	3.6822	2.6593	0.110575	增设5个平台	3.6071	1.8426	0.048671

从表2可以看出，增设4个平台时效果较好，最大出警时间变化不大，工作量标准差也比较小，具体的增设位置为29, 40, 48和87号路口．

3 全市交巡警服务平台的相关问题

这部分针对全市主城六区的交巡警服务平台需要解决下面两个问题：

1) 要求评价该市六区设置平台方案的合理性，如果明显不合理，则给出解决方案；

2) 该市地点P发生了重大刑事案件，要求给出调度全市平台警力资源的最佳围堵方案．

3.1 全市平台设置方案的合理性评价模型

该问题可以从两个方面考虑，一是全市六区的平台不考虑分区限制，即将全市六区视为一个整体的网络，可以统一调度管理；二是全市六区的平台分区调度管理，即分别考虑六个区的具体情况．在这里只考虑第一种情况，第二种情况与问题1)的方法相类似．

对于全市平台设置合理性的主要指标有最大出警时间

$$T=\max_{1\le j\le n}(t_j),$$

其中$t_j\,(j=1,2,\cdots,n)$表示平台y_j管辖区域内的最大出警时间，相应的工作量为$G_j\,(j=1,2,\cdots,n)$，其均值为$\overline{G}$，最大工作量为

$$G=\max_{1\le j\le n}(G_j),$$

则各平台的工作量标准差为

$$\sigma=\sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(G_j-\overline{G}\right)^2}.$$

根据问题附件中所给全市的相关数据和A区平台管辖范围的分配模型，用启发式算法求解，则可以得到全市平台管辖范围的分配方案．从全市情况来看，在全市的582个路口节点中总共有138个是不能在3分钟内到达的，按就近原则分配这些路口节点．将其它的路口节点在到达时间不超过3分钟的约束条件下寻求最优的分配方案，不难得到全市的80个平台的平均最大出警时间为3.7735分钟，最大的出警时间为12.6803分种；平均每天的出警量为8.4312次，最多的出警量为26.1次，最少为1.6次．由此可知，现有全市的平台设置还不尽合理，需要再增设一些新的平台．

3.2 增设新平台方案的确定模型

交巡警平台设置的合理性主要体现在两个方面：各平台的最大出警时间尽量小和总工作量尽量均衡．解决该问题的方法与A区的解决方法不同．因为路口数量多，不能对所有路口进行全搜索实现，要根据现有平台设置明显不合理的地方进行增设新平台．因此，在这里定义一个与平台的出警时间和工作量有关的指标，称为“需求度”，作为判断在一个路口是否需要增设平台的依据，从而找出最需要增设平台的路口．对于每个路口 x_i 定义需求度如下

$$\mu_i = \lambda \frac{t_i}{t_{\max}} + (1-\lambda)\frac{G_i}{G_{\max}},$$

其中

$$t_i = \sum_{j=1}^{n} T_{ij} = \sum_{j=1}^{n} B_{ij} X_{ij}$$

为路口 x_i 到其所管辖平台的距离；

$$t_{\max} = \max_{\substack{1\le i\le m \\ 1\le j\le n}} T_{ij}$$

为所有出警时间的最大值；G_i 为管辖路口 x_i 平台的工作量，

$$G_{\max} = \max_{1\le i\le m} (G_i)$$

为所有平台最大工作量；λ 为权重．对于每一个没有设平台的路口计算出相应的需求度值，如果需求度较大的路口比较集中在一个区域，则说明在该区域现有平台的设置有明显的不合理，从而应在该区域内增设新平台来降低需求度．根据上述的方法，对比增设1个至6个平台的情况，如表3所示．

表3: 增设新平台的对应结果比较分析

增设平台数	增设平台编号	最大出警时间(分钟)	最大工作量(次/天)	工作量标准差
0	未增设	12.6803	26.1	4.4946
1	314	12.6803	18.4	4.1365
2	517	12.6803	18.4	3.9251
3	388	9.8992	18.4	3.9753
4	330	8.1188	18.4	3.9973
5	288	8.1188	15.9	3.8714
6	206	8.1188	15.9	3.8794

从表3可以看出，前5个平台增设后都对三个指标产生明显的改善效果，但是增加第6个平台后却没有对指标产生明显的变化．故在全市范围内增设5个平台就可以改善现有平台设置明显不合理的状况．

3.3 全市范围内最佳围堵方案的确定模型

将事发地点P到其它各路口节点的最短路程(或时间)用矩阵 $D=(D_1,D_2,\cdots,D_m)$ 表示．不妨设交巡警和犯罪嫌疑人逃跑的车辆速度均为60 km/h．在实际中，犯罪嫌疑人逃跑的时间不同，可能逃跑的区域范围也不同．其逃跑的区域范围可以看作以事发地P为中心沿交通网络向外辐射的一个区域．在时间 t 内将逃跑区域内所有路口节点的集合记为 $F(t)$，即为嫌疑人逃跑 t 时间后可能逃跑的区域范围，则有 $F(t)=\{x_i\,|\,D_i\le t,1\le i\le m\}$．要将犯罪嫌疑人封锁在 $F(t)$ 范围内，$F(t)$ 的所有边界路口节点集合记为 $E(t)$，即 $E(t)$ 中的任一节点都属于 $F(t)$，且与 $F(t)$ 之外的邻近节点邻接，故有

$$E(t)=\{x_i\,|\,x_i\in F(t),\,x_j\notin F(t),\,l_{ij}<\infty,1\le i\le m,1\le j\le n\},$$

其中 $x_j\notin F(t)$ 为 $x_i\in F(t)$ 的区域外邻接点，l_{ij} 为邻接距离．

对于任一个路口节点 $x_i\in E(t)$，需要封锁的时间为 D_i-3 分钟．因此要全封锁区域 $F(t)$，即 $E(t)$ 内所有路口节点需要的最短时间为 $\min\limits_{x_i\in E(t)} D_i-3$ 分钟，最长时间为 $\max\limits_{x_i\in E(t)} D_i-3$ 分钟．

在全市范围内最优的封锁方案应该使得封锁区域尽量小，而封锁区域的大小可以由集合 $F(t)$ 的秩(节点个数)来表示，即 $Q=\mathrm{Rank}(F(t))$．要保证封锁范围的有效性，即确保封锁住的必要条件为 $B_{ij}X_{ij}(t)\le D_i(t)-3,\ i=1,2,\cdots,p,\ j=1,2,\cdots,n$，其中 $X_{ij}(t)=0$ 或 1，即表示 t 时间要平台 y_i 去封锁路口 x_i 取1，否则取0，$p=\mathrm{Rank}(E(t))$．则有最佳的围堵优化模型：

$$\min Q=\mathrm{Rank}\big(F(t)\big),$$

$$\text{s.t.}\begin{cases}F(t)=\{x_i\,|\,D_i\le t,1\le i\le m\},\\ E(t)=\{x_i\,|\,x_i\in F(t),\,x_j\notin F(t),\,l_{ij}<\infty,1\le i\le m,1\le j\le n\},\\ B_{ij}X_{ij}(t)\le D_i(t)-3,\quad i=1,2,\cdots,p,\quad j=1,2,\cdots,n,\\ p=\mathrm{Rank}\big(E(t)\big).\end{cases}$$

这里决策变量 $X_{ij}(t)$ 由下面的优化模型确定：

$$\min Y=\max_{\substack{1\le i\le p\\1\le j\le n}} B_{ij}X_{ij},$$

$$\text{s.t.}\begin{cases}\sum\limits_{i=1}^{p}X_{ij}\le 1,\quad j=1,2,\cdots,n,\\ \sum\limits_{j=1}^{n}X_{ij}=1,\quad x_i\in E(t),\quad i=1,2,\cdots,p,\\ \sum\limits_{j=1}^{n}X_{ij}=0,\quad x_i\notin E(t),\quad i=1,2,\cdots,p,\\ x_{ij}=0\text{或}1,\quad i=1,2,\cdots,p,\quad j=1,2,\cdots,n.\end{cases}$$

对于上面的复杂的优化模型采用启发式搜索算法，可以得到一种有效的全封锁方案，具体结果如表4所示．

表4: 一种有效的全封锁方案和对应封锁时间 (时间单位: 分钟)

平台编号	封锁节点	封锁时间	最短逃跑时间	平台编号	封锁节点	封锁时间	最短逃跑时间
A16 (16)	16	0	3.3016	C5 (170)	243	5.0978	8.3009
A5 (5)	5	0	3.8768	F1 (475)	561	4.3548	8.7969
A6 (6)	6	0	3.9074	A4 (4)	60	1.7392	8.8017
C8 (173)	235	0.5315	4.1282	C4 (169)	240	7.0474	10.1541
A15 (15)	15	0	4.1386	A1 (1)	41	4.4412	10.5025
A3 (3)	55	1.2659	5.2104	C3 (168)	4	3.6994	10.7557
C6 (171)	232	1.9064	5.7845	C17 (182)	273	2.1024	12.8088
A10 (10)	10	0	6.1882	C10 (175)	168	4.9779	13.3739
C7 (172)	244	3.5289	6.8121	D1 (320)	370	7.8085	16.7721
A2 (2)	3	2.1117	7.1106	D2 (321)	371	8.9798	17.1079
A17 (17)	40	2.6879	7.9625	C2 (167)	248	3.6788	17.5237

从表4可以看出，当封锁时间小于嫌疑人逃出封锁区域的最短时间减去3分钟时，即交巡警能在嫌疑人之前到达要封锁的目标路口，则实现了对应路口的有效封锁．需要对22个路口进行封锁，最长的封锁时间为8.9798分钟，即调度全市的交巡警平台的警力，在9分钟之内可以完成有效全封锁．从交巡警要封锁的时间和嫌疑人逃出区域的时间来看，该封锁方案是有效的，即能确保将嫌疑犯被封锁在该区域范围之内．

4 关于竞赛论文的综合评述

交巡警服务平台的设置与调度问题作为2011年“高教社杯”全国大学数学建模竞赛本科组的B题，根据在全国和部分赛区的论文统计情况，大约有三分之一的本科组参赛队选择了这个题目．由于这个题目的特点和解决难度，我们在赛前估计选做这个问题的人会少是预料之中的，但与选做A题数量差距如此之大是出乎预料之外的，这成为命题人的一个遗憾．赛后，我们与部分指导教师交流和评阅专家分析，出现这种情况的主要原因有两点：一是与该问题密切相关的参考资料几乎没有；二是解决这个问题对模型表述、模型的求解算法和编程能力要求都比较高，这让很多同学望而却步．

从另一个侧面也反映出来，我们的学生对文献资料的依赖性过强，而针对实际问题的独立思考、自主创新意识却不够强；同时，对实际数据的处理能力、算法设计和应用编程能力也不够强，总体反映出数学建模的能力不足．因此，如何加强培养学生独立思考和自主创新意识，提高数学建模能力，都是值得广大的数学建模指导教师思考和在培训中注意的问题．

尽管如此，我们也高兴地看到，很多参赛队在三天的时间里非常好地解决了这个复杂的问题，而且有的参赛队给出很好的解决思路和方法，都是很值得赞赏的．譬如：武汉大学的一个参赛队所建立的全市范围的围堵模型很有特色，建模思路清晰、模型表述简洁、自己设计的启发式法巧妙，程序实现也很有创新之处，为此获得了全国一等奖，但在文字表述和符号的使用上也有不足之处．上海财经大学一个参赛队的围堵模型也有不错的表现．另外，被选择发表的几篇优秀论文都有各自不同的特点，多数都是基本问题完成的比较好，论文表述也比较清晰，但对全市的围堵模型普遍完成的不太好(具体可参考他们发表的论文)．当然，论文是学生在三天的时间内完成的，存在一些差错和不足均属正常．

4.1　解决问题有代表性的方法概述

第一部分，针对A区的情况主要包括三个小问题，首先需要根据题目所提供的数据，计算出交通网络节点间的邻接矩阵和最短路矩阵等，这也是解决问题的基础，然后按题目要求分别解决三个问题.

1)　针对为A区20个平台合理分配管辖范围的问题，首先要确定合理性的指标，通常包括平台的出警时间尽量在3分钟之内和工作量尽量均衡；然后以每一个平台在所管辖范围内的最大出警时间最小和各平台的工作量的方差(或标准差、绝对差等)为目标，建立20个平台管辖范围的最优分配模型，即可归结为一个不完全的指派模型. 可以直接求解双目标的优化模型，也可以根据实际情况简化为单目标优化模型来求解. 可以用Lingo直接实现，也可以用Matlab编程实现求解.

2)　针对A区13个出入口的全封锁问题，以最长的封锁时间最小为目标，以一个平台最多封锁一个路口和每一个路口有且仅有一个平台封锁为约束，建立问题的指派模型，直接用Lingo求解即可得到最优的封锁方案.

3)　针对在A区增加2至5个平台的方案问题，首先确定一个需要增加平台的原则和判断增加平台位置的准则，判别在哪里最需要增加平台. 逐个增加2至5个平台后，再按问题1)的方法重新分配各平台的管辖范围，计算各平台的最大出警时间和出警工作量的均衡性指标. 根据增加平台后的效果，确定出需增加平台的个数和位置. 同时要说明需要增加k个，而不需增加$k+1\,(2\leq k\leq 4)$个的理由.

第二部分，针对全市六区的情况主要包括两个小问题：分析评价全市现有平台设置的合理性和确定最佳的围堵方案.

1)　针对分析评价全市六区现有平台设置的合理性问题，可以在两种不同的假设下考虑问题. 一种是将全市六区视为一个整体的网络图，不考虑各区的界限. 全市80个平台可以跨区管辖邻近路口节点，参照A区平台管辖范围分配方法，即可得到的管辖范围分配方案是一个整体的评价方案. 事实上，在全市的582个路口节点中共有138个是不可能在3分钟内到达的路口节点，按就近原则分配这些路口节点. 将其它的路口节点在到达时间不超过3分钟的约束条件下，按各平台的工作量尽量均衡的原则就近分配管辖范围. 另一种是将全市六区分区管理，即分别视为六个独立的交通网络图，采用A区分配管辖范围的方法，分别给出各区的分配方案和合理性的评价结果. 判断合理性的指标为各平台的最大出警时间和工作量的均衡性. 对于明显不合理的地方增加新平台的方法与A区的情况类似.

2)　针对确定全市范围内的围堵方案问题，首先要确定在这段时间内嫌疑人可能逃跑的最大区域，针对这个区域找出需要封锁的目标路口节点. 再利用对A区全封锁问题的解决方法，合理调度全市的80个平台的警力(不受区的限制)，对可能逃跑区域范围的目标路口进行全封锁即可.

事实上，选择封锁的区域越大，区域内所包含的路口节点就越多，实际要进行搜捕的工作量会越大；实现封锁时间越长，犯罪嫌疑人逃跑的范围可能就越大；如果选择封锁的区域越小，实现搜索的时间就越短，但犯罪嫌疑人可能会跑出封锁区域. 我们既要求合理地选择封锁范围，从而保证安全封锁，而同时又使得所需要的封锁时间尽量少. 既要实现全封锁的时间最少，又要封锁的区域最小，即区域内的路口节点数最少.

要求确定封锁方案，就是以封锁区域最小、也可以需要封锁区域内的路口节点数最小为目标，建立实现全封锁方案的优化模型. 注意到，需要封锁区域的大小是随时间变化的，能封锁住的基本约束条件是：

$$\text{出事地点到将要封锁的路口所需时间} + 3\text{分钟} \geq \text{指派平台到封锁路口的所需时间}.$$

对于问题的求解需要设计启发式算法编程实现，具体的实现需要的一定的技巧和方法.

4.2 竞赛论文存在的主要问题概述

纵览送全国和部分赛区评审的参赛论文，普遍存在的问题主要有以下几个方面.

4.2.1 A区相关问题的情况

1) 针对A区为交巡警服务平台合理分配管辖范围的问题，很多论文都是依就近分配的原则来分配管辖路口节点，即以最短路或最短出警时间为目标，将72个路口节点分配给20个平台. 但没有从实际出发适当考虑各平台出警工作量的均衡性问题，导致了分配结果明显的不合理，甚至有几个平台只管平台所在的一个路口(典型的有6号、10号和14号平台)，有的平台管的路口很多，工作量严重的失衡. 虽然，问题中没有明确要求考虑工作的均衡，但从数学建模解决实际问题的角度，应该要考虑这个问题，这也反映出参赛学生对数学建模创新性的认识理解还不够，更多的还是基于应用题层面上就事论事地回答问题.

2) 在考虑对A区的13个路口实施紧急封锁的问题时，很多论文都是以封锁完成的总时间或平均时间最小为目标，建立指派模型，求解0-1规划模型得到封锁方案，这都是不太合理的. 解决该问题的核心是用最短时间封锁所有的路口，所以应该取最大的封锁时间最小化，相应的时间才是完成全封锁的时间.

3) 在考虑增设平台的问题时，很多论文只考虑3分钟内到达的要求目标，即只针对现有6个3分钟内不能到达的路口，通过直观分析就确定增设4个平台的结果，而没有兼顾综合考虑各平台出警工作量不均衡的因素. 同时没有通过求解优化模型给出增设$k\,(1\leq k\leq 5)$个平台，而不需要增设$k+1$个平台的理由，即分析比较增设2至5平台的最优选择.

4.2.2 全市六区相关问题的情况

1) 对于全市六区现有平台设置的合理性分析问题，很多论文把这个问题看成一个综合评价问题，用了简单的层次分析法、综合加权法、TOPSIS方法等，也简单地得到一个合理与不合理的结论. 凡此种论文都没有充分地说明合理与不合理的理由，更没有给出解决不合理的具体方法. 如增设平台，增设多少个和增设的具体位置在哪里，增设以后的效果怎么样，如何重新分配管辖范围等问题.

2) 对于全市范围内交通路口围堵的问题，绝大多数的论文都没有给出可行的围堵方案. 很多人把问题看成一个确定性的问题，有的直观分析选择要封锁的路口，就近确定平台执行封锁；也有的建立一个确定的指派模型求解得到一个确定的方案. 甚至，有不少论文没有考虑到保证封锁住的基本条件，所给出的方案没有办法保证能封锁的住. 还有个别的论文，简单地直接去封锁出市区17个出口，没有考虑后续的搜捕问题.

3) 很多论文所给出的模型没有详尽的分析与解释，让别人很难看懂. 有些仅有模型没有具体的求解算法，只是说用软件求解得到结果；有的也没有详细的程序，难易判断其模型和结果的正确性. 有的论文所给出的结果只是一些数据，或图表，没有任何分析与说明，让人很难判断其有效性和正确性.

4.3 进一步研究的问题

虽然2011年全国大学生数学建模竞赛已经结束，但从我们评阅的论文来看，对于交巡警服务平台的设置与调度问题无论是从实际问题本身，还是从相关问题理论、模型和算法等方面，仍然还存在一些有待于进一步研究的问题. 在这里提出来，希望有兴趣的同学和老师做进一步深入的研究.

1) 关于平台的管辖范围合理分配问题，是否能用区域覆盖的方法，给出一个连续区域分配方案，而不是仅考虑路口节点.

2) 如果全市范围内的交巡警平台可以统一调度，没有区的界限，则能否给出全市所有平

台管辖范围的合理分配方案？要有相应的数学模型和相应有效的求解算法与结果.

3) 从全市角度出发，不考虑区的界限，根据现有平台的设置情况，从出警时间尽量短和各平台的出警工作尽量均衡两个方面考虑，通过建模分析提出增加平台的具体方案，包括具体的数量和位置.

4) 关于全市范围“动态”围堵数学模型简明准确的表述方法和求解算法的进一步研究.

5) 关于全市范围内围堵的最好方案究竟是什么？需要围堵区域是什么？包括哪些路口节点？现有正确的最好围堵区域结果包含21个到22个路口节点.

6) 实现全市范围内的围堵最快的时间是多少？现有正确最好的结果是8.9分钟，是否还有更快的可行方案？

7) 能否证明目前的结果已是最好的结果？

8) 根据现有围堵方案所给出的区域范围，进一步研究在这个区域内最有效的搜捕模型.

对于上述问题的进一步研究都是有意义的，研究结果需要给出具体、正确详尽的数学模型、求解算法、明确的结果和可靠性分析，必要情况需要给出理论证明和程序代码.

5 结束语

2011年“高教社杯”全国大学生数学建模竞赛工作圆满结束了，也正是在迎来这项赛事20周年之际，我们能为竞赛活动尽微薄之力，有幸成为今年的命题人，倍感荣幸. 我们也高兴地看到，2011年的竞赛规模达到了19490个队，并且多个国家和地区的几十个境外参赛队参赛，即我们的竞赛正式走出了国门，走向了世界，这也充分说明了这项活动的生命力和数学建模事业的号召力. 在此，我们衷心地祝愿这项竞赛活动更健康有序地发展，进一步提高我国建模水平和应用水平，努力推动数学建模事业的发展，宏扬数学建模的精神，为社会培养出更多高素质的有用人才.

致谢：在题目形成和加工过程中，得到了全国组委会、各位专家的指导和帮助，特别是周义仓、姜启源、孟大志和边馥萍四位教授多次在一起深入细致地讨论问题，谭永基教授和谢金星教授等也都给予了很多的帮助，还有重庆市交巡警总队有关领导给予了很大的支持，在这里一并表示衷心的感谢！

参考文献：

[1] 全国大学生数学建模竞赛组委会. 2011年竞赛题[OL]. http://www.mcm.edu.cn, 2011
Organizing Committee of the National Mathematical Contest in Modeling. 2011 CUMCM Problems[OL]. http://www.mcm.edu.cn, 2011

[2] 韩中庚. 数学建模方法及其应用(第2版)[M]. 北京：高等教育出版社，2009
Han Z G. Methods and Applications of Mathematical Modeling (2nd Edition)[M]. Beijing: Higher Education Press, 2009

[3] 赵静，但琦. 数学建模与数学实验(第3版)[M]. 北京：高等教育出版社，2008
Zhao J, Dan Q. Mathematical Modeling and Mathematical Experimenats (3rd Edition)[M]. Beijing: Higher Education Press, 2008

Optimized Model of Setting and Scheduling of Traffic and Patrol Police Service Platforms

DAN Qi[1], HAN Zhong-geng[2], YANG Ting-hong[1]

(1- Department of Basic, Logistics Engineering University, PLA, Chongqing 401311;
2- Institute of Information Engineering, Information Engineering University, PLA, Zhengzhou 450002)

Abstract: In this paper, according to the problem B of 2011 Higher Education Press Cup CUMCM, the background of the problem of setting and scheduling of traffic and patrol police service platforms is introduced. Then the approaches and the concrete models of several related problems are given, as well as the main solving methods and the corresponding results, and the general existing problems of the participating papers are analyzed briefly. Finally, several problems worthy of further study are given.

Keywords: traffic and patrol police service platforms; the transportation network map; model of distribution; surrounds model

第28卷 增刊1
2011年11月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 28 Supp.1
Nov. 2011

文章编号:1005-3085(2011)07-0117-08

交巡警服务平台的设置与调度

胡立忠， 张金磊， 朱金宝

指导教师： 李安平， 安芹力， 梁放驰

(空军工程大学导弹学院，西安 713800)

编者按: 该文思路清晰，推理严谨，表达清楚，特别对图论和优化模型的运用恰当，对问题的分析较为透彻．在建模过程中，能抓住问题的关键，并利用Matlab与Lingo软件给出了具体的结果．不足之处是对问题二中的围堵方案没有进一步讨论．下面的论文中略去了问题的重述、模型假设和符号说明等内容．

摘 要: 根据城市的实际情况与需求合理性来设置交巡警服务平台、分配各平台的管辖范围、调度警务资源是警务部门的一个实际课题．本文采用Floyd算法，得到A区中距离各路口最近的交巡警服务平台，给出服务平台管辖范围的分配方案．以封锁时间最短和总出警路程最小为目标函数建立优化模型，得到封锁方案．以出警时间不超过3分钟，选择使工作量方差最小和增设平台数量最少为目标，得到增设平台数量．针对现有设置方案，发现138个点发生案件时交巡警3分钟内无法到达．以出警时间不超过3分钟，建立以增设平台以及重新布局两种模型对全市6个区的警备资源配置进行调整．在犯罪嫌疑人围堵问题中，以围堵时间最小、围堵区域面积最小和人口最少为目标得到围堵方案．

关键词: 0-1规划；Floyd算法；最短路问题；均衡度

分类号: AMS(2000) 90C29; 90B20　　**中图分类号**: O221　　**文献标识码**: A

1 问题分析

1.1 问题一的分析

1) 要为各交巡警服务平台分配管辖范围，首先需要利用图论中的相关算法求得任意平台间的最短路径[1]；其次可考虑依据就近原则将路口分配给距离最近的交巡警平台管辖，这样就可最大限度满足管辖范围的分配要求．

2) 要从20个交巡警服务平台选择13个实现道路快速封锁，合理的调度方案应满足该方案中最后一个到达交通要道路口的交巡警服务平台所用时间最少．如需要进一步考虑该方案的优化，则还应满足出警总路程达到最短．

3) 增加平台时可将3分钟内到达案发现场这一要求作为约束条件，求得不满足要求的路口，为使这些路口满足条件，应在这些路口附近(包括本身)进行增设；考虑到工作量的均衡，可将工作量的方差作为衡量均衡度的指标，进而确定增加平台的个数与位置．

1.2 问题二的分析

1) 本问题要求按照设置交巡警服务平台的原则和任务，分析研究该市现有交巡警服务平台设置方案的合理性．因此，建立模型时应综合考虑最大出警时间和工作量的均衡性这两个因素，在所得结果的基础上再对不合理的平台提出改进方案．

2) 由于在案发后，犯罪嫌疑人会立即逃窜，且其逃窜路线是随机的，因此，在建立模型时，考虑的是将“出事地点到将要封锁的路口所需时间大于等于指派平台到封锁路口的所需时间加3分钟”作为能封锁住的基本约束条件，并在这个约束条件之下给出最优封锁方案．

2 模型的建立与求解

在建立模型前，首先将警车实际行驶速度 $v = 60\,\text{km/h}$ 转换为图上行驶速度 v'，由地图比例尺计算公式得

$$v' = \frac{v}{100000} = 10\,\text{mm/min},$$

即出警时交巡警每走一分钟对应图上10mm的路程．其次，利用图论中Floyd算法通过Matlab软件[2,3] 求出任意平台与任意路口间的最短距离．

2.1 问题一模型的建立与求解

2.1.1 管辖范围分配问题

为满足尽量快速到达事发地的要求，采用就近原则将各路口分配给路程最近的服务平台管辖．记第 i 个交巡警平台距第 j 个交通要道路径为 l_{ij}，建立目标函数为

$$f_j = \min_{i=1}^{20} l_{ij}, \quad j = 1, 2, \cdots, 92.$$

具体A区交巡警平台管辖范围的分配方案，如表1所示．

表1: A区交巡警平台管辖范围的分配方案

平台	管辖范围	平台	管辖范围	平台	管辖范围
1	1, 67, 68, 69, 71, 73, 74, 75, 76, 78	2	2, 39, 40, 43, 44, 70, 72	3	3, 54, 55, 65, 66
4	4, 57, 60, 62, 63, 64	5	5, 49, 50, 51, 52, 53, 56, 58, 59	6	6
7	7, 30, 32, 47, 48, 61	8	8, 33, 46	9	9, 31, 34, 35, 45
10	10	11	11, 26, 27	12	12, 25
13	13, 21, 22, 23, 24	14	14	15	15, 28, 29
16	16, 37, 38	17	17, 41, 42	18	18, 80, 81, 82, 83
19	19, 77, 79	20	20, 84, 85, 86, 87, 88, 89, 90, 91, 92		

需要说明的是，通过计算发现A区中编号分别为28, 29, 38, 39, 61, 92的6个路口交巡警无法在3分钟内到达．距离最近交巡警服务平台到达这6个路口的最短时间，如表2所示．依据就近原则，这6个路口只能分配给距它们最近的平台管辖．

表2: 最近平台到达6个交叉路口的最短时间(单位：分钟)

平台	路口	最短时间	平台	路口	最短时间	平台	路口	最短时间
16	38	3.40588	7	61	4.1902	2	39	3.68219
20	92	3.60127	15	28	4.75184	15	29	5.70053

2.1.2 警力合理调度问题模型的建立与求解

1) 最快全封锁模型的建立与求解

基于问题分析，可将题意表述如下：寻找一种方案，在20个交巡警服务平台中选择13个，对13条交通要道进行封锁，使封锁时间最小．为满足快速实现全封锁，采用以下步骤：

a) 在20个交巡警服务平台中选择13个，对13条交通要道进行封锁，得到一个可行方案；

b) 每种可行方案对应13条出警路线，封锁时间取决于其中最长出警路线；

c) 对这些最长出警路线进行比较，选择其中最小的一个作为调度方案．

由此可得模型如下

$$\min\ e=\max_{1\le j\le 13}\left\{\sum_{i=1}^{20}e_{ij}\right\}$$

$$\begin{cases}\sum\limits_{i=1}^{20}a_{ij}=1,\\ \sum\limits_{j=1}^{13}a_{ij}\le 1,\\ e_{ij}=a_{ij}l_{ij},\end{cases}$$

其中

$$a_{ij}=\begin{cases}1,&\text{第}i\text{个平台封锁第}j\text{个交通要道},\\ 0,&\text{第}i\text{个平台不封锁第}j\text{个交通要道},\end{cases}$$

l_{ij} 表示第 i 个交巡警平台到第 j 个交通要道的最短路程；$e_{ij}=a_{ij}l_{ij}$ 表示完成对 j 交通要道封锁的出警路线路程．

约束条件[4] 为

$$\sum_{i=1}^{20}a_{ij}=1,\quad j=1,2,\cdots,13,\quad \text{表示一个路口必须由一个交巡警平台来实施封锁；}$$

$$\sum_{j=1}^{13}a_{ij}\le 1,\quad i=1,2,\cdots,20,\quad \text{表示一个交巡警平台最多只封锁一个路口．}$$

利用Lingo软件编程 $e=80.15\,\mathrm{mm}$，即交巡警服务平台最快可在8.015分钟内对13条交通要道进行全封锁，最佳调度方案如表3所示．在本方案中，交巡警出警所走总路程为461.88mm．

表3: 最佳调度方案

调度方案	调度路径	调度方案	调度路径
$2\to 38$	$2\to 40\to 39\to 38$	$4\to 62$	$4\to 62$
$5\to 48$	$5\to 47\to 48$	$7\to 29$	$7\to 30\to 29$
$8\to 30$	$8\to 33\to 32\to 7\to 30$	$9\to 16$	$9\to 35\to 36\to 16$
$10\to 22$	$10\to 26\to 11\to 22$	$11\to 24$	$11\to 25\to 24$
$12\to 12$	$12\to 12$	$13\to 23$	$13\to 23$
$14\to 21$	$14\to 21$	$15\to 28$	$15\to 28$
$16\to 14$	$16\to 14$		

2.1.3 增加平台模型(工作量均衡模型)的建立与求解

将3分钟内有交巡警到达事发地作为增加交巡警平台的约束条件，即$l_{ij} \leq 30\,\mathrm{mm}$.

交巡警一次出警的工作量可用其出警途中耗费时间T和处理案发现场耗费时间T'之和来表述，那么一个服务平台的工作量即为：$W_i = \sum (T_i + T'_i)p_i$. 工作量均衡用其方差[5]表示，得目标函数

$$\min\ S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \overline{W}\right)^2.$$

综上可建立如下模型

$$\min\ S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \overline{W}\right)^2$$

$$\begin{cases} W_i = \sum (T_i + T'_i)p_i, \\ l_{ij} \leq 30, \\ 2 \leq n \leq 5. \end{cases}$$

在现有交巡警服务平台设置情况中由2.1.1可知有6个路口不符合约束条件，因此增加的平台必须存在于这6个路口周围，进而由约束条件寻找可能增设平台的位置. 通过分析，至少需要分别从$\{28, 29\}$，$\{38, 39, 40\}$，$\{48, 61\}$，$\{87, 88, 89, 90, 91, 92\}$这4组路口集合中选择4个路口增设才能满足设置标准. 根据已建立的模型，运用Matlab软件优化结果，得到满足要求的平台位置为28, 39, 48, 87这四个路口. 此时其工作量方差最小，为5.6185. 如在此基础上再增设一个平台来进一步优化分配管辖范围，做法与寻找4个点的情况相同，最后得到一个最优的增设平台位置为69. 相对4个增设平台位置的情况，再增设一个平台的工作量最小方差为4.2925，优化效果不明显. 另外，考虑到警备资源、经济状况等实际问题，所以无需增设5个交巡警服务平台. 因此在28, 39, 48, 87这四个路口增设新的交巡警服务平台为较合理的方案.

由此得到增设平台后交巡警服务平台的管辖范围，如表4所示.

表4: 增设平台后管辖范围分配方案

平台	管辖范围	平台	管辖范围	平台	管辖范围
1	1, 67, 68, 69, 71, 73, 74, 75, 76, 78	2	2, 43, 44, 70, 72	3	3, 54, 55, 65, 66
4	4, 57, 60, 62, 63, 64	5	5, 49, 50, 51, 52, 53, 56, 58, 59	6	6
7	7, 30, 32	8	8, 33, 46	9	9, 31, 34, 35, 45
10	10	11	11, 26, 27	12	12, 25
13	13, 21, 22, 23, 24	14	14	15	15, 28, 29
16	16, 36, 37	17	17, 41, 42	18	18, 80, 81, 82, 83
19	19, 77, 79	20	20, 85, 86	28	28, 29
39	38, 39	48	47, 48, 61	87	84, 87, 88, 89, 90, 91, 92

2.2 问题二模型的建立与求解

2.2.1 现有平台设置合理性分析

为分析现有平台设置的合理性，用Floyd算法求得全市内所有路口到达最近服务平台的最短路程．发现存在多达138个路口发生案件时3分钟内交巡警无法到达，这严重不符合处理案件的及时性这一要求．考虑以3分钟内有交巡警可以到达案发现场为约束条件、以设置平台数最少为目标函数，在现有平台设置的基础上增设服务平台(方案一)或不考虑现有平台重新设置服务平台(方案二)两种方案来求得需要设置的平台数．

同时为了让每个服务平台的工作量均衡，用2.1.3中工作量均衡模型对平台设置的具体位置及管辖范围求解．

2.2.2 需增加平台数模型建立

基于2.2.1分析，新增平台后应该满足所有路口发生案件时3分钟内有交巡警可以到达案发现场．

模型一　在不改变现有平台布局的情况下增设平台，使所有路口都能在3分钟内有巡警到达，并使各平台工作量尽量均衡．建立如下模型

$$\min\ \sum_{i=1}^{m} p_i - n$$

$$\begin{cases} l_{ij}x_{ij} \le 30, \quad i=1,2,\cdots,m, \quad j=1,2,\cdots,m, & \text{(a)}\\ \sum\limits_{i=1}^{n} x_{ij} = 1, \quad j=1,2,\cdots,m, & \text{(b)}\\ x_{ij} \le p_i, \quad i=1,2,\cdots,m, \quad j=1,2,\cdots,m, & \text{(c)}\\ x_{ij} = 0 \text{ 或 } 1, \quad i=1,2,\cdots,m, & \text{(d)}\\ p_i = 0 \text{ 或 } 1, \quad i=1,2,\cdots,m, & \text{(e)}\\ p_i = 1, \quad i=1,2,\cdots,n, & \text{(f)} \end{cases}$$

其中n为已有平台数，考虑到警力资源的有限性，将增加平台数$\sum\limits_{i=1}^{m} p_i - n$作为目标函数，这里

$$p_i = \begin{cases} 1, & \text{第}i\text{个路口是平台},\\ 0, & \text{第}i\text{个路口不是平台}. \end{cases}$$

m为总路口数，l_{ij}第i个平台到第j个路口的最短路程

$$x_{ij} = \begin{cases} 1, & \text{第}i\text{个平台管辖第}j\text{个路口},\\ 0, & \text{第}i\text{个平台不管辖第}j\text{个路口}. \end{cases}$$

(a)式表示每个平台管辖的所有点能够3分钟赶到；(b)式表示每个路口只能由一个平台管辖；(c)式表示只有在第i个路口为服务平台时才能将第j个路口让其管辖；(f)式表示前n个路口为已有平台．

此模型确定了增加平台的个数，而具体增设位置及管辖范围应考虑到要尽量使平台的工作量均衡，用5.1.3中的工作量均衡模型求解．

用Lingo编程求得增设平台情况，如表5所示．

表5: 解决方案一

区域	新增平台个数	新增平台位置
A	4	28, 39, 48, 87
B	2	104, 149
C	15	184, 199, 201, 204, 205, 216, 238, 240, 252, 257, 259, 263, 288, 304, 318
D	8	329, 332, 333, 338, 344, 362, 369, 371
E	14	387, 389, 390, 392, 393, 408, 419, 420, 421, 446, 454, 458, 472, 474
F	13	491, 509, 513, 517, 539, 541, 558, 560, 569, 574, 575, 578, 582

模型二 不考虑已有平台，对所有路口进行重新布局，使所有路口都能在3分钟内有交巡警到达，并使各平台工作量尽量均衡. 类似模型一，可建立如下模型

$$\min \sum_{i=1}^{m} p_i$$

$$\begin{cases} l_{ij}x_{ij} \le 30, \quad i=1,2,\cdots,m, \quad j=1,2,\cdots,m, \\ \sum\limits_{i=1}^{n} x_{ij}=1, \quad j=1,2,\cdots,m, \\ x_{ij} \le p_i, \quad i=1,2,\cdots,m, \quad j=1,2,\cdots,m, \\ x_{ij}=0 \text{ 或 } 1, \quad i=1,2,\cdots,m, \\ p_i=0 \text{ 或 } 1, \quad i=1,2,\cdots,m. \end{cases}$$

用Lingo编程求得增设平台情况，如表6所示.

表6: 解决方案二

区域	新增平台个数	新增平台位置
A	15	3, 10, 11, 14, 15, 22, 25, 29, 33, 40, 43, 48, 58, 66, 90
B	5	104, 114, 135, 149, 163
C	27	166, 175, 177, 179, 184, 185, 193, 199, 201, 204, 205, 214, 222, 231, 238, 240, 250, 254, 256, 259, 263, 266, 274, 285, 303, 310, 319
D	12	327, 329, 332, 333, 338, 344, 350, 361, 362, 363, 369, 371
E	23	373, 378, 387, 389, 390, 393, 397, 400, 408, 416, 419, 420, 421, 423, 424, 440, 444, 446, 457, 459, 470, 472, 473
F	19	479, 485, 488, 489, 499, 509, 511, 515, 532, 539, 541, 559, 562, 565, 570, 573, 574, 575, 582

比较模型一与模型二，发现模型二只需设101个平台，而模型一需增加56个，即共设136个平台. 从警力资源的有限性考虑，选择模型二结果作为调整方案.

2.2.3　最佳围堵问题模型的建立与求解

此问题可类似于2.1.2中的警力合理调度问题，不同的是需要调度全市警力围堵封锁，而完成围堵封锁的最短时间必须小于嫌疑人逃窜时间减去3分钟，即

$$\left\{\min\ e=\left[\max_{1\le j\le n}\left(\sum_{i=1}^{80}e_{ij}\right)\right]\right\}\le t-3.$$

因此可建立模型如下：

目标函数为

$$\min t,$$

约束条件为

$$\begin{cases}\left\{\min\ e=\left[\max\limits_{1\le j\le n}\left(\sum\limits_{i=1}^{80}e_{ij}\right)\right]\right\}\le t-3,\\ vt<\overline{r_{pj}},\\ e_{ij}=a_{ij}l_{ij},\\ a_{ij}=0\text{ 或 }1,\\ \sum\limits_{i=1}^{80}a_{ij}=1,\\ \sum\limits_{j=1}^{n}a_{ij}\le 1.\end{cases}$$

利用Lingo编程求解可得最佳围堵方案，如表7所示．在此方案中，接到报案后7.36min就可形成包围圈，满足快速搜捕要求，为最佳围堵方案．

表7:　围堵具体方案

序号	围堵路径	序号	围堵路径	序号	围堵路径
1	13, 22, 21, 14	2	17, 17	3	12, 27, 26
4	15, 28, 29	5	2, 43, 42, 27, 41	6	3, 44, 2, 43
7	5, 50, 51, 59, 60, 62	8	1, 69, 68	9	10, 34, 35, 45, 44, 2, 70
10	19, 77, 76	11	4, 62, 190, 189, 168	12	173, 232, 231, 216, 215
13	171, 228, 218	14	172, 227	15	169, 254, 253, 240
16	167, 249, 248	17	174, 224, 223, 225, 170, 273	18	320, 349, 371
19	481, 491, 486, 487, 489, 482,	20	477, 498, 495, 493, 492, 491, 486, 487	21	476, 545, 546, 547, 532 548, 549
22	478, 554, 555, 557, 558	23	475, 557, 564, 563, 562		

3 模型的评价

1) 建立的模型简洁明了，使用了数学工具Matlab、Lingo等软件，提高了工作效率.

2) 修改模型参数可满足各种同类需求，具有很强通用性.

当然模型还存在许多需要改进的地方，例如可考虑车速变化等问题.

参考文献：

[1] 吴祈宗. 运筹学与最优化方法[M]. 北京：机械工业出版社，2003
Wu Q Z. Operations Research and the Optimization Method[M]. Beijing: Mechanical Industry Press, 2003

[2] 宋叶志，贾东用. Matlab数值分析与应用[M]. 北京：机械工业出版社，2009
Song Y Z, Jia D Y. Numerical Analysis and Application of Matlab[M]. Beijing: Mechanical Industry Press, 2009

[3] 林雪松，周婧，林德新. Matlab7.0应用集锦[M]. 北京：机械工业出版社，2006
Lin X S, Zhou Q, Lin D X. Matlab7.0 Application[M]. Beijing: Mechanical Industry Press, 2006

[4] 姜启源，谢金星，叶俊. 数学模型[M]. 北京：高等教育出版社，2004
Jiang Q Y, Xie J X, Ye J. Mathematical Model[M]. Beijing: Higher Education Press, 2004

[5] 韩中庚. 数学建模方法及其应用[M]. 北京：高等教育出版社，2005
Han Z G. Mathematical Modeling and Application[M]. Beijing: Higher Education Press, 2005

Setting and Scheduling of Traffic and Patrol Police Service Platforms

HU Li-zhong, ZHANG Jin-lei, ZHU Jin-bao

Advisor: LI An-ping, AN Qin-li, LIANG Fang-chi

(Missile Institute, Air Force Engineering University, Xi'an 713800)

Abstract: It is an important issue to allocate and schedule the traffic and patrol police service platforms (TPPSPs) in accordance with the practical police force and condition. By employing the Floyd algorithm, the distances between the locales and TPPSPs are obtained and the jurisdiction region for each TPPSP in the city center District A is allocated. Based on the object function of minimizing the blockade time and the shortcut, the blockade scheme for the TPPSPs of District A is scheduled. The supplemental TPPSPs' number is accessed on condition that the police can arrive at the locale within 3 minutes with the minimum workload variance of the TPPSPs of District A and the minimum number of supplemental TPPSPs. Through investigating the reasonableness of the TPPSP setting for the entire city, the 138 access arteries cannot arrive at the locale within 3 minutes. The models of added and reallocated TPPSPs are adjusted on condition that the police can arrive at the locale within 3 minutes. An optimal containment strategy is established according to the object function of minimum blockade time, region area and region population.

Keywords: 0-1 program; Floyd algorithm; short-path problem; balance degree

第28卷 增刊1
2011年11月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 28 Supp.1
Nov. 2011

文章编号:1005-3085(2011)07-0125-10

基于0-1规划的交巡警平台设置与调度模型

唐 棣[1]，董小小[1]，魏 歆[1]

指导教师： 罗万春[2]

(1- 第三军医大学学员旅九队，重庆 400038； 2- 第三军医大学数学教研室，重庆 400038)

编者按: 本文利用Floyd算法、0-1规划、指派模型较好地解决了交巡警平台的设置、管辖区域的划分以及发生重大突发事件警务资源的调度问题．本文特色是较好地利用工作量的变异系数，分析研究工作量的均衡性，为合理设置调整交巡警平台位置提供了定量依据．

摘 要: 本文对交巡警平台的设置与调度问题进行了研究．首先，根据Floyd算法求得平台与路口间的最短距离，使各路口被与之距离最近的平台管辖．发生突发事件时，以最快到达指定路口并且总调度距离最短为目标，基于0-1规划，建立了平台警力的调度模型．然后，以各平台工作量的变异系数最小和最长出警时间最短为目标，重新对平台管辖区域进行划分，并得出了新增平台的方案．最后，为抓捕从P点逃跑的犯罪嫌疑人，我们以嫌疑犯可能逃窜的最长时间最短为目标，以交巡警成功封锁路口和嫌疑犯被完全围堵为约束条件，通过0-1规划得到了最终的围堵方案：调度18个平台的警力封锁18个路口．按照最优围堵方案，最多20.52分钟即可拦截住嫌疑犯．

关键词: 最短路径；0-1规划；交巡警平台

分类号: AMS(2000) 93A30　　**中图分类号**: U491.8　　**文献标识码**: A

1 模型假设

1) 交巡警出警时间是指从交巡警平台到达事发地路口节点所用的时间；
2) 交巡警平台管辖区域的划分对象为路口节点；
3) 一般情况下，各个交巡警平台的管辖范围相互独立；
4) 警车的平均时速为60 km/h；
5) 全封锁是以最后一个路口节点完成封锁为标志；
6) 常规情形下，全市各区的交巡警平台不跨区管理；
7) 每个路口仅由一个平台管辖，每个平台可管辖多个路口；
8) 嫌疑犯的平均逃跑速度与警车的平均速度相同．

2 符号说明

m: 研究范围内路口的个数；

n: 研究范围内交巡警平台的个数；

l: 研究范围内进出口个数；

S_{ij}: 交巡警平台 j 到路口 i 的距离；

V: 警车时速；

C_i: 路口 i 的案发率；

W_j: 交巡警平台 j 的工作量，即平台 j 管辖范围内各路口案发率的总和；

T_j: 第 j 个平台的最长出警时间．

3 模型的建立与求解

3.1 问题一：A 区交巡警平台的设置与调度分析

3.1.1 A 区交巡警平台的管辖范围分配

当出现突发事件时，为使交巡警警力尽量在3分钟内 (警车的时速为60 km/h) 到达事发地点，需要各路口由与之距离最近的交巡警平台来管辖．首先可根据Floyd算法[1] 计算出平台与各路口间路径之和最小值．再结合图1中的流程图，利用Matlab编程[2]，可找出距离各路口最近的平台并计算出其距离．

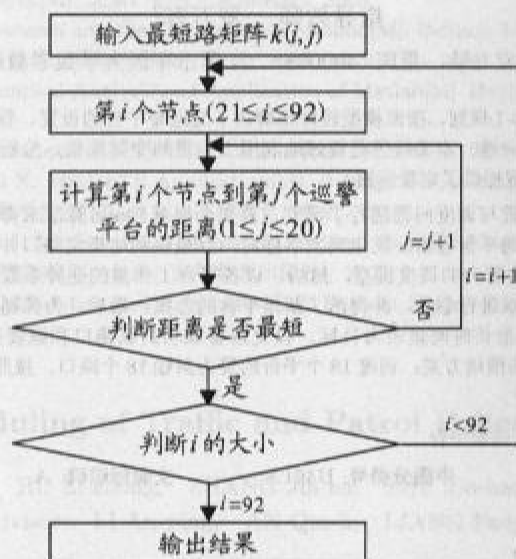


图1: A区寻找距离各路口最近的交巡警平台的流程图

由此可得各平台的管辖范围，见表1.

表1: 各平台的管辖范围

交巡警平台	节点	交巡警平台	节点
A1	1 67 68 69 71 73 74 75 76 78	A11	11 26 27
A2	2 39 40 43 44 70 72	A12	12 25
A3	3 54 55 65 66	A13	13 21 22 23 24
A4	4 57 60 62 63 64	A14	14
A5	5 49 50 51 52 53 56 58 59	A15	15 28 29
A6	6	A16	16 36 37 38
A7	7 30 32 47 48 61	A17	17 41 42
A8	8 33 46	A18	18 80 81 82 83
A9	9 31 34 35 45	A19	19 77 79
A10	10	A20	20 84 85 86 87 88 89 90 91 92

3.1.2　A区13条交通要道的快速封锁调度方案

发生重大突发事件后，调度20个交巡警服务平台的警力资源，对进出该区的13条交通要道实现快速全封锁，实际中一个平台的警力最多封锁一个路口，因此这属于指派问题.

对于各平台，只有调度和不调度两种情况，因此，可用0-1规划的思想建立模型[3]. 设 x_{kj} 为第 k 个出口被第 j 个平台的警力封锁的情况，则有

$$x_{kj}=\begin{cases}0,\\ 1,\end{cases}\quad k=1,2,\cdots,l,\quad j=1,2,\cdots,n, \tag{1}$$

其中 $x_{kj}=0$ 表示第 k 个出口不被第 j 个平台的警力封锁，$x_{kj}=1$ 表示第 k 个出口被第 j 个平台的警力封锁，n 表示研究范围内交巡警平台的个数，l 表示研究范围内出口个数.

为了在最短时间内实现全封锁，而全封锁的时间是由封锁最后一个路口所用的时间决定的. 因此，目标函数就是求最远调度距离的最小值，表述为

$$\min Z=\max\Big(\sum_{j=1}^{n}S_{kj}x_{kj}\Big), \tag{2}$$

其中 Z 表示所有调度中的最远距离，S_{kj} 表示第 j 个平台到第 k 个出口的距离，该距离可由Floyd算法得出.

约束条件为：

1)　平台不一定被调去封锁出口，即

$$\sum_{k=1}^{l}x_{kj}\le 1,\quad j=1,2,\cdots,n. \tag{3}$$

2)　出口只被一个平台封锁，则有

$$\sum_{j=1}^{n}x_{kj}=1,\quad k=1,2,\cdots,l. \tag{4}$$

综上，最快实现全封锁的模型为[4]：

$$\min Z=\max\Big(\sum_{j=1}^{n}S_{kj}x_{kj}\Big),$$

$$\text{s.t.}\begin{cases}\sum\limits_{k=1}^{l}x_{kj}\le 1,\quad j=1,2,\cdots,n,\\ \sum\limits_{j=1}^{k}x_{kj}=1,\quad k=1,2,\cdots,l,\\ x_{kj}\in\{0,1\}.\end{cases} \tag{5}$$

根据模型(5)，利用Matlab编程和题目中给出的相关数据，最后可以得到若干组解. 这些解的共同特点是最远的调度距离都是平台7至出口9的距离，为80.15百米. 所以完成A区全封锁的最短时间即由此决定，需要8分钟.

为了得到最优的调度方案，在此基础上，增加“最远的调度距离是平台7至出口9的距离”这一约束条件，以总调度距离最短为目标函数，对除平台7和出口9以外的出口和交巡警平台

进一步作0-1规划，模型为

$$\min Y=\sum_{k=1}^{l-1}\sum_{j=1}^{n-1}S_{kj}x_{kj},$$

$$\text{s.t.}\begin{cases}\max\left(\sum\limits_{j=1}^{n}S_{kj}x_{kj}\right)\leq S_{97},\\ \sum\limits_{k=1}^{l}x_{kj}\leq 1,\quad j=1,2,\cdots,n-1,\\ \sum\limits_{j=1}^{k}x_{kj}=1,\quad k=1,2,\cdots,l-1,\\ x_{kj}\in\{0,1\},\end{cases}\tag{6}$$

其中Y表示除平台7到出口9距离以外的总调度距离，$(n-1)$表示除平台7以外的平台总数，$(l-1)$表示除出口9以外的出口总数.

利用Lingo软件编程求解[5]，最终的调度方案见表2.

表2: 最快实现完全封锁且总距离相对最短的调度方案

出口	1	2	3	4	5	6	7	8	9	10	11	12	13
平台	A12	A16	A8	A14	A10	A13	A11	A15	A7	A9	A2	A5	A4
距离(百米)	0.00	67.42	26.92	32.65	77.08	5.00	38.05	47.52	80.15	34.92	39.82	24.76	3.50

综上，最快实现完全封锁的时间为8分钟，调度的总距离为477.79百米.

3.1.3 增加平台后管辖范围的分配方案

由于各平台管辖范围内的路口数差异很大，以及各路口的案发率不同，造成现有交巡警服务平台的工作量不均衡，部分地方的出警时间过长．首先，在不增加交巡警服务平台的前提下重新分配管辖范围.

对于路口i有被平台j管辖和不被平台j管辖两种情况，因此，可设计0-1变量．令

$$x_{ij}=\begin{cases}0,\\ 1,\end{cases}\quad i=1,2,\cdots,m,\quad j=1,2,\cdots,n,\tag{7}$$

其中$x_{ij}=0$表示第i个路口不被第j个平台管辖，$x_{ij}=1$表示第i个路口被第j个平台管辖，n表示研究范围内交巡警平台的个数，m表示研究范围内路口的个数.

管辖区域的划分原则为使各平台的工作量尽量均衡，出警时间尽量缩短.

目标函数1 用各平台工作量的变异系数来表示工作量的均衡性，其值越小表示工作量越均衡．首先，各交巡警平台的工作量可表示为该平台管辖范围内所有路口案发率的总和，即

$$W_j=\sum_{i=1}^{m}C_ix_{ij},\tag{8}$$

其中W_j指交巡警平台j的工作量，C_i表示路口i的日案发率.

所有平台工作量的平均值为

$$\overline{W}=\frac{1}{n}\sum_{j=1}^{n}W_j,\tag{9}$$

则各平台工作量的变异系数最小表示为

$$\min f_1 = \frac{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(W_j-\overline{W})^2}}{\overline{W}}. \tag{10}$$

目标函数2　所有平台的最长出警时间相比，其中的最大值最小表示最长出警时间最短．交巡警的出警时间是指从接警到到达事发地路口的时间，则第 j 个平台的最长出警时间为

$$T_j = \max\left(\frac{S_{ij}x_{ij}}{V}\right), \tag{11}$$

其中 S_{ij} 表示第 j 个平台到达第 i 个路口的最短距离．则最长出警时间最短表示为

$$\min f_2 = \max(T_j). \tag{12}$$

约束条件如下：

1)　平台不闲的约束．为使每个平台不至于无管辖范围，可约束为它至少管辖自己所在的路口．当 $i=j$ 时，即

$$x_{ij}=1, \quad i=j. \tag{13}$$

2)　每个路口都被平台管辖的约束．当 $i\neq j$ 时，第 i 个路口必定被 $1\sim n$ 中的唯一一个平台管辖，即

$$\sum_{j=1}^{n} x_{ij}=1, \quad i\neq j. \tag{14}$$

3)　出警时间不超过3 min，即

$$\frac{S_{ij}}{V}\leq 3. \tag{15}$$

综上，考虑平台的工作量均衡及出警时间合理的模型[6]为

$$\min f_1 = \frac{\sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(W_j-\overline{W})^2}}{\overline{W}}, \quad \min f_2 = \max(T_j),$$

$$\text{s.t.}\begin{cases} x_{ij}=1, \quad i=j, \\ \sum_{j=1}^{n} x_{ij}=1, \quad i\neq j, \\ \frac{S_{ij}}{V}\leq 3, \\ x_{ij}\in\{0,1\}. \end{cases} \tag{16}$$

在不增加交巡警平台的前提下，根据模型(16)，利用Matlab软件进行求解，结果见表3．据此可知在此前提下，最长出警时间为5.70 min，出现在平台15前往路口29处理突发事件时．工作量的变异系数为0.1830．

为缩短最长出警时间，将新增平台设置在出警时间最长的路口处．同理可求得增加平台1到5个时工作量变异系数及最长出警时间的变化，见表4．

结论：增加4个交巡警平台时，分别位于路口28(29)、61、39和91，此时最长出警时间已达到3 min以内的要求，为2.71 min，工作量的变异系数较小，为0.2004．增加5个交巡警平台

时，分别位于路口28(29)、61、39、91和67，工作量的变异系数减小，为0.1526，最长出警时间仍为2.71 min．因此，增加4或5个交巡警平台都满足题目要求．考虑到建设成本时，可只增加4个交巡警平台；若更重视工作量的均衡性，需增加5个交巡警平台．

表3：最长出警时间最短且工作量均衡时各平台的管辖范围

交巡警平台	管辖的路口	日工作量(案件数)
A1	1 71 73 74 75 68	6.5
A2	2 43 44 70 69	6.9
A3	3 54 55 65 66 67	6.4
A4	4 57 60 62 63 64	6.6
A5	5 49 52 53 56 58	6.9
A6	6 50 59 47 51 48	6.9
A7	7 30 61	5.1
A8	8 33 46 32	6.5
A9	9 31 35 45	6.5
A10	10 34	3.7
A11	11 26 27	5.6
A12	12 25 24	5.1
A13	13 22 23	6.0
A14	14 21	4.9
A15	15 28 29	4.8
A16	16 36 37 38 39	6.4
A17	17 41 42 40 72	6.8
A18	18 81 82 83 84 90 86	8.4
A19	19 76 77 78 79 80	6.1
A20	20 87 88 89 91 92 85	8.4

表4：增加平台后工作量的变异系数和最长出警时间

新增平台个数	新增平台位置	工作量的标准差	工作量的均值	变异系数	最长出警时间
0	无	1.14	6.23	0.1830	5.70
1	28或29	1.34	5.93	0.2260	4.19
2	61	1.51	5.66	0.2668	3.82
3	39	1.35	5.41	0.2495	3.68
4	91	1.04	5.19	0.2004	2.71
5	67	0.76	4.98	0.1526	2.71

3.2 问题二：全市交巡警平台的设置与调度

3.2.1 全市现有交巡警平台设置的合理性分析及调整方案

首先，在不增加交巡警平台时，根据问题一中的模型(16)对主城六区分别进行分析，结果见表5.

表5: 全市现有交巡警平台的相关数据

主城六区	路口数	平台数	平均工作量	变异系数	最长出警时间
A区	92	20	6.23	0.1830	5.70
B区	73	8	6.36	0.1743	4.47
C区	154	17	8.85	0.1725	6.86
D区	52	9	5.24	0.2070	16.06
E区	103	15	5.79	0.1979	19.10
F区	108	11	5.60	0.1883	8.48

根据表5中的数据，我们对全市(A，B，C，D，E，F)现有的交巡警服务平台设置方案的合理性进行分析，结果发现：

1) 主城各区工作量的变异系数都较小，即各交巡警平台的工作量均衡，比较合理；

2) 主城各区的最长出警时间都较大，尤其是D区和E区，远远超过了规定的出警时间3 min，不合理，因此各区的交巡警服务平台都有待调整．造成这一结果的原因主要是部分路口与最近平台间的距离超过3 km.

针对全市各区交巡警平台的出警时间过长这一问题，我们在不改变原有交巡警平台位置的基础上，增加尽量少的平台，使最长出警时间小于3 min.

利用问题一中A区增加平台的方法，可分别求出B、C、D、E、F各区新增平台的个数及位置，并进一步求解得增加平台后各区交巡警平台的最长出警时间，结果见表6.

表6: 增加平台位置及增加前后各区最长出警时间的比较(单位：分钟)

主城六区	不增加平台时	新增平台(路口编号)	增加平台后
A区	5.70	28 39 61 92	2.71
B区	4.47	123 152	2.91
C区	6.86	166 167 168 169 170 171 174 175 176	
		177 178 179 180 183 199 201 203 205	2.99
D区	16.06	320 322 324 325 326 328 329	2.91
E区	19.10	372 373 374 376 378 379 380 381 382	
		383 384 385 386 387 388 390 393	2.97
F区	8.48	475 477 478 479 480 482 483 484 485 486	
		490 505	2.95

由表6可知，新增平台后，主城六区的最长出警时间与不增加平台时相比，明显缩短，全部控制在3 min之内，因此更合理.

3.2.2 最佳围堵方案

3.2.2.1 模型的建立

以最快抓捕嫌疑犯为目标，表述为

$$\min U = \max\left(\frac{L(i,32)}{V}\right), \tag{17}$$

其中$L(i,32)$表示从事发地(路口32)到落网地点(路口i)的距离，U表示嫌疑犯从逃跑到落网的时间，根据假设8，嫌疑犯的逃跑速度为V.

约束条件为：

1) 由于部分路口的封堵对抓捕嫌疑犯无意义，因此，不一定每个路口都被封堵，即

$$\sum_{j=1}^{n} x_{ij} \le 1, \quad i = 1,2,\cdots,m, \tag{18}$$

$$x_{ij} = \begin{cases} 0, \\ 1, \end{cases} \quad i = 1,2,\cdots,m, \quad j = 1,2,\cdots,n, \tag{19}$$

其中$x_{ij}=0$表示第i个路口不被第j个平台封堵，$x_{ij}=1$表示第i个路口被第j个平台封堵，n表示研究范围内交巡警平台的个数，m表示研究范围内路口的个数.

2) 不一定每个平台都出警，即

$$\sum_{i=1}^{m} x_{ij} \le 1, \quad j = 1,2,\cdots,n. \tag{20}$$

3) 嫌疑犯到达路口i的时间减3 min比交巡警从平台j到路口i的时间长，表示路口i能成功的被平台j封堵，即

$$\frac{L(i,j)}{V} \le \frac{L(i,32)}{V} - 3, \tag{21}$$

其中$L(i,j)$表示平台j到路口i的距离.

4) 为描述嫌疑犯被封堵在一定的区域内，定义路口i被封堵为$L'(Q,i)=\inf$，即其它任意路口与路口i不相通，距离表示为inf. 若Q_1表示被封堵区域以外所有路口的集合，Q_2表示被封堵区域内所有路口的集合，则

$$L'(Q_1,Q_2) = \inf. \tag{22}$$

综上，模型为

$$\min U = \max\left(\frac{L(i,32)}{V}\right),$$

$$\text{s.t.} \begin{cases} \sum\limits_{j=1}^{n} x_{ij} \le 1, \quad i = 1,2,\cdots,m, \\ \sum\limits_{i=1}^{m} x_{ij} \le 1, \quad j = 1,2,\cdots,n, \\ \frac{L(i,j)}{V} \le \frac{L(i,32)}{V} - 3, \\ L'(Q_1,Q_2) = \inf, \\ x_{ij} \in \{0,1\}. \end{cases} \tag{23}$$

3.2.2.2 模型的求解

将相关数据代入模型(23)，可得到最终的封堵方案，见表7.

表7: 全市各区的围堵方案

调度平台	封锁路口	逃跑时间	封锁时间	调度平台	封锁路口	逃跑时间	封锁时间	调度平台	封锁路口	逃跑时间	封锁时间
A1	63	8.63	3.50	A16	16	3.30	0.00	C4	240	10.15	6.94
A2	3	6.48	2.11	A17	40	7.96	2.69	C6	245	5.61	2.57
A3	55	5.21	1.27	A18	41	10.50	5.54	C7	231	6.96	2.78
A4	4	8.80	0.00	A19	62	9.13	5.03	C8	246	6.54	3.08
A10	10	6.19	0.00	C2	248	20.52	3.67	D1	371	15.89	7.36
A15	29	9.16	5.70	C3	168	12.59	0.00	F1	561	8.79	4.35

完成以上所有路口的封锁后，即可成功围堵嫌疑犯．完成全区域封锁的时间为7.36分钟，嫌疑犯最迟会在20.52分钟被成功拦截．

参考文献：

[1] 叶奇明，石世光. Floyd算法的演示模型研究[J]. 海南大学学报(自然科学版)，2008, 26(1): 47-50
Ye Q M, Shi S G. The research of Floyd algorithm demo model[J]. Journal of Hainan University (Natural Science), 2008, 26(1): 47-50

[2] 马莉. Matlab数学实验与建模[M]. 北京：清华大学出版社，2010
Ma L. Matlab Mathematics Experiments and Modeling[M]. Beijing: Tsinghua University Press, 2010

[3] 朱茵，王军利，周彤梅. 智能交通系统导论[M]. 北京：中国人民公安大学出版社，2007
Zhu Y, Wang J L, Zhou T M. Introduction to Intelligent Transport System[M]. Beijing: Chinese People's Public Security University Press, 2007

[4] 姜启源，谢金星，叶俊. 数学模型(第3版)[M]. 北京：高等教育出版社，2003
Jiang Q Y, Xie J X, Ye J. Mathematical Modeling (3rd Edition)[M]. Beijing: Higher Education Press, 2003

[5] 谢金星. 优化建模与Lindo/Lingo软件[M]. 北京：清华大学出版社，2011
Xie J X. Optimization Modeling and Lindo/Lingo Software[M]. Beijing: Tsinghua University Press, 2011

[6] 张锦，王坤. 流线网络优化的变分不等式模型与算法[J]. 西南交通大学学报，2011, 46(3): 481-487
Zhang J, Wang K. Variational inequality model and algorithm for stream line network optimization[J]. Journal of Southwest Jiaotong University, 2011, 46(3): 481-487

The Setting and Scheduling Model of Traffic and Patrol Police Platform Based on 0-1 Programming

TANG Di[1], DONG Xiao-xiao[1], WEI Xin[1]
Advisor: LUO Wan-chun[2]

(1- The 9th Team of Cadet Brigade, Third Military Medical University, Chongqing 400038;
2- Department of Mathematics, Third Military Medical University, Chongqing 400038)

Abstract: This paper aims at resolving the problem of the setting and scheduling of the traffic and patrol police platform. First, the shortest path and distance between the intersection and platform is calculated based on the Floyd algorithm, and each intersection is set to be dominated by the nearest platform. In emergency case, the model is established on the basis of 0-1 programming to attain the fastest and the shortest scheduling scheme. By compromising the workload of each platform and the time wasted on the way to the intersection into consideration, we establish a partition model for the territorial jurisdiction and achieve the scheme of setting new platforms. Finally, in order to catch the criminal fled from Point P as soon as possible, we construct the optimization model by setting the likely longest feeing time to be shortest as the objective and that the police can successfully close the intersections and the criminal can be completely blocked up as the constraint conditions. The best plan for blockade can then be drawn by solving the model using 0-1 programming. The final result is as follows: the police from 18 platforms should be scheduled to close 18 intersections and the criminal is to be caught within 20.52 minutes.

Keywords: shortest path; 0-1 integer programming; the police platform

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0135-10

交巡警服务平台的设置与调度

邢中阳， 曾凡正， 张江彬
(国防科技大学光电科学与工程学院，长沙 410073)

编者按：本文在研究交巡警平台管辖区域划分时，能较详细的分析交巡警平台出警次数，工作量均衡．在此基础上构建多目标规划模型，计算出最长封锁时间，并评价了新增交巡警平台位置的合理性．本文特色是采用设置两个包围圈的方法，给出了最佳围堵方案、算法和最优结果．

摘 要：本文以出警时间和平台工作量均衡为目标，构建了交巡警平台管辖区的划分模型及交巡警平台增加的优化模型，并利用Floyd算法求出最短路，再通过调整算法均衡各平台间工作量，给出A区交巡警平台管辖区域划分结果，并确定了新增平台位置分别为28、36、48、57、81．为实现对A区13条交通要道的快速全封锁，构建了以最长封锁时间最短和平均封锁耗时最短为两级目标的封锁模型，利用二分-匈牙利算法求得最短封锁时间为8.015分钟．在全市交巡警平台设置方案评价模型中，提出了综合考虑工作量期望、方差及最长出警时间的评价函数F，并将其用于交巡警平台的调整．最后讨论了考虑内外封锁圈的最佳围堵方案．

关键词：最短路；3分钟覆盖；交巡警平台

分类号：AMS(2000) 90C30 中图分类号：O221.2 文献标识码：A

1 模型假设

1) 假设题中所给的报案节点即为实际案发位置，各路段不发生拥堵，道路都是双行道；
2) 认为两节点的道路是直线，但本文方法也适用于道路弯曲的情况．

2 问题分析和符号说明 (略)

3 模型的建立与求解

3.1 A区内交巡警平台管辖区域划分模型

对于其管辖区域的划分问题，本质上是一个以最大的出警时间最短为目标，同时兼顾工作量均衡的一个覆盖问题[1,2]．通过下面两步完成目标：

第1步 最短路的方案：求出距离每个节点最近的交巡警平台．设离第i个节点最近的交巡警平台为M_i，其最短距离为D_{iM_i}，警车速度为v．假设每个工作平台警力不限，则问题即为求$z=\min\max D_{iM_i}/v$，即到达所有节点的最长出警时间最短．

第2步 在此基础上采用调整算法，实现各平台间工作量均衡．于是可在最短路方案的基础上提出改进方案，把工作量也作为一个目标考虑进去．

3.1.1 最短路方案下的辖区分配方案

最短路方案是基于题中要求“3分钟内尽量有交警到达案发地”为目标的，找出距离每个节点最近的交巡警服务平台，在此方案下求出的节点与平台间的路径总是最短的，在最短的情况如果不能实现“3分钟覆盖”，那么其它的方案也无法实现．根据Floyd算法[3,4]计算结果，如表1所示．

分析表1数据，发现最短路算法下的工作量极不均衡，因此提出优化方案–最短路–工作量方案.

表1: 最短路径下A的辖区分配

平台号	辖区内的节点号	工作量	平台号	辖区内的节点号	工作量
1	1 67 68 69 71 73 74 75 76 78	10.3	11	11 26 27	4.6
2	2 39 40 43 44 70 72	9.7	12	12 25	4
3	3 54 55 65 66	5.6	13	13 21 22 23 24	8.5
4	4 57 60 62 63 64	6.6	14	14	2.5
5	5 49 50 51 52 53 56 58 59	9.7	15	15 28 29	4.8
6	6	2.5	16	16 36 37 38	5
7	7 30 32 47 48 61	9.6	17	17 41 42	5.3
8	8 33 46	5	18	18 80 81 82 83	6.1
9	9 31 34 35 45	8.2	19	19 77 79	3.4
10	10	1.6	20	20 84 85 86 87 88 89 90 91 92	11.5

3.1.2 考虑工作量均衡下的最短路辖区划分方案的进一步优化

由于改进方案的调整本身就是一个粗调，因此在筛选出工作量较大的平台时，对筛选标准进行简化，以平台管辖区内的节点个数为标准，筛选出工作量过大的几个平台.

对于某个平台，选出距离平台最远的几个节点，将这几个节点与周边其他平台进行匹配，匹配的条件是：满足“3分钟覆盖”以及目标平台的节点数为饱和．如果有满足匹配条件的平台，那么选择距离最近的平台，将管辖权转移；否则，节点仍归属原平台.

利用新方案中的算法，对各平台间的管辖权进行调整，可以得出调整后的辖区分配表，如表2所示.

表2: 调整后的辖区分配

平台号	辖区内的节点号	工作量	平台号	辖区内的节点号	工作量
1	1 69 71 74	5	11	11 26 27	4.6
2	2 40 43 44 72 70 68	9.2	12	12 25	4
3	3 54 55 65 66 67	6.4	13	13 22 23 24	7.1
4	4 57 60 62 63 64	6.6	14	14 21	3.9
5	49 50 51 52 53 56	7.7	15	15 28 29 31	6.4
6	6 47 58 59	6.1	16	16 36 37 38 39	6.4
7	7 30 48 61	6.1	17	17 41 42 92	6.1
8	8 33 46 32	6.5	18	18 81 82 83 88 73 90	8
9	9 34 35 45	6.6	19	19 77 79 80 76 78 75	7.9
10	10	1.6	20	20 85 86 87 89 91	7.9

3.1.3　工作量均衡前后的对比分析

均衡前后各平台工作量对比情况，如图 1 所示．调整前工作量方差

$$\sigma^2 = \sum_{i=1}^{20} (a_i - \mu)^2 = 8.009875,$$

均衡后工作量方差

$$\sigma^2 = \sum_{i=1}^{20} (a_i - \mu)^2 = 2.846475.$$

后者在工作量均衡度方面明显比前者更优．在最终方案中得到管辖区域划分的最终结果直观表示图，如图 2 所示．

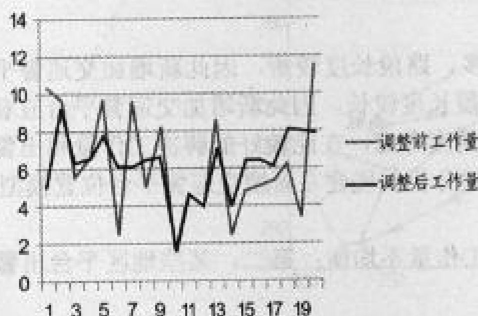


图 1:　调整前后各平台工作量对比

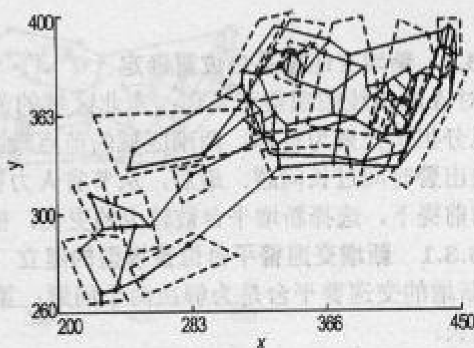


图 2:　最终辖区划分图

3.2　封锁方案的求解

以“最小化最大的最短距离”为标准，利用二分枚举和使用最优性剪枝的搜索算法求得最优封锁路线．其中“最小化最大的最短距离”是指在所有封锁指派方案中，13 个需要封锁的路口与对应平台需要的出警时间的最大值在所有方案中最小．

最直接的方法是用枚举法，将每种封锁方案搜索出来，筛选出最优方案．对于获得的多组最优方案的选择，采用多级目标[5]：

I 级目标—最长封锁时间最短：$\min\max D_{imap_i}$；

II 级目标—封锁平均耗时：$\min\sum D_{imap_i}$；

这里 map_i 为某个指派方案中，第 i 个要封锁的路口对应分配的交巡警平台的编号．

若枚举全部状态，则由排列组合知识可知有 A_{20}^{13} 种方案，求每种匹配方案中的距离采用 Floyd 算法求出的最短距离．

进一步优化可构造一种更加简洁的算法—二分-匈牙利算法．基于二分法的一些性质，可以来查找满足条件的临界点．查找的下界为 $l = 0$；查找的上界为给定区域中最远两点距离，不妨令其为原图对角线的长度 $r = \sqrt{500^2 + 600^2} \approx 782$．显然，对于 r，必定可以实现全封锁．但仅靠上述算法不能求出具体的分配方案，接下来使用最优性剪枝搜索方法可以把所有符合条件的方案找出来，从中选择满足 II 级目标的最优解．得到最终封锁模型示意图，如图 3 所示．计算求得最长封锁时间为 8.015 分钟，为由 7 号平台处的警力去封锁 29 号节点这一段．

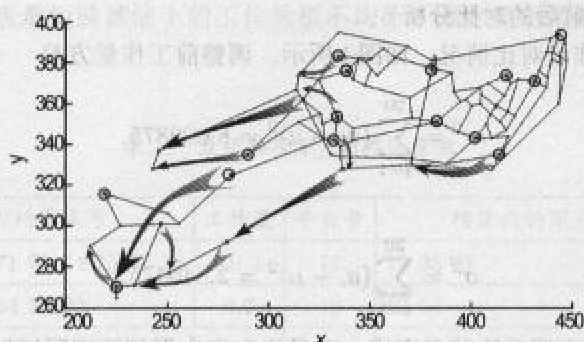


图 3:　封锁模型示意图

3.3　新增交巡警平台位置确定

对A区路线图观察可发现：东北区域的节点较多，路段长度较短，因此新增加交巡警平台应以分担工作量为目标；西南区域的节点较少，路段长度较长，因此新增加交巡警平台应着力解决出警时间过长问题．最后，从节省人力资源的角度考虑，在能较好的解决工作量和出警时间的前提下，选择新增平台数较少的更优．根据上述分析，现建立新增交巡警平台位置模型．

3.3.1　新增交巡警平台位置模型的建立

新增的交巡警平台是为解决两个问题：第一，工作量不均衡；第二，某些地区平台出警时间过长．

出警时间的衡量标准是：无法在3分钟内到达距离自己最近的平台的节点，都可以认为是"过远"．将这样的路段选出，在路段中点处插入节点，交巡警平台可以设置在该节点处，以起到缩短出警时间的作用．

假设现有交巡警平台数量为M，且位置确定，路口节点数量为N．

设各路段长度为l_i，对于$l_i/v \geq 6\,\mathrm{min}$的路段，在路段中点处插入节点，设新插入节点处的案发率为0，这样插入的总节点数为$n(n=1,2,\cdots)$．新插入的节点与原来节点的地位完全相等，总节点数为$N+n$．设新增交巡警平台的数量为m．

现在问题变为：如何从$N+n-M$个节点中选出m个设置平台，并对现有$M+m$个平台的管辖区域进行适当划分，使得警力分配资源更优．

定义工作量方差为σ^2，最短的最长出警时间为z，那么问题可以描述为$\min\{g(\sigma^2,z)\}$．

抽象的变量为位置$i=0,1,2,\cdots,n$，其中

$$\sigma^2=\sum\frac{(\text{工作量}-\text{平均值})^2}{M+m},\quad z=\min\ \max\frac{D_{iM_i}}{v}.$$

$g(\sigma^2,z)$为工作量方差和最长出警时间的一个评价函数．这是一个双目标的优化问题，两个无相关性的目标并行，第一目标为：$\min\sigma^2$，第二目标为：$z\leq 3\,\mathrm{min}$．

3.3.2　新增交巡警平台数量及位置的确定

由于工作量方差和最长最短路这两个指标相关性不是很大，因此对其分别进行考虑．

题中对新增平台数m的约束为：$2\leq m\leq 5$，因此对$m=0,1,\cdots,5$的情况分别进行讨论，其中$m=0$仅作为一个原始参考数据，$m=1$作为辅助参考数据．

在计算了工作量方差和最短的最长出警时间两方面数据后，综合两个目标，可以求得目标函数的最优解，即为了解决工作量分配不均及某些地方出警时间过长问题，本文选择增加五个

平台，位置分别是28, 36, 48, 57, 81．对现有的25个平台重划分管辖区域，得到新辖区内的出警情况及工作方差，如表3所示．新增平台的直观示意图，如图4所示．

表3：新增平台前后的各关键数据比对

	出警时间超过3分钟的个数	最长出警时间 (min)	平均出警时间 (min)	工作量方差
$m=0$	6	5.700525	1.122	3.980875
$m=5$	0	2.7083140	0.844	1.073243

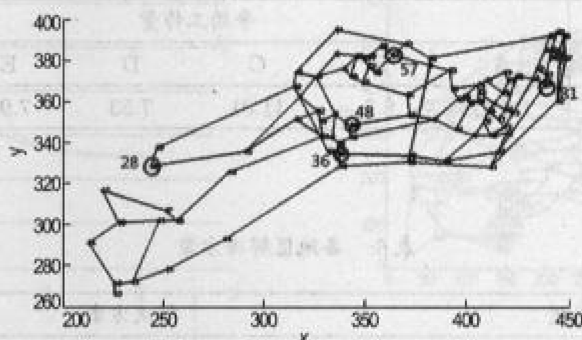


图4：新增平台示意图

3.4 全市交巡警平台设置方案的评价模型

设置交巡警服务平台的主要原则是：能及时有效地赶到案发点，发生重大案件时能及时对附近路口进行封锁，各个交巡警平台间工作量相对均衡．

全市的交巡警平台分布是A区平台分布的一个推广，可以应用前面的模型对管辖区域做适当划分后再利用工作量方差和出警时间作为评价指标，衡量平台设置的合理性．但直接选取城市这个整体为对象，所需计算量过大，因此我们分别对各区建模讨论，最终将各区数据整合，得到对全城的评价方案．

3.4.1 分别对各城区建立评价模型

分析评价的指标为：最长出警时间、无法实现3分钟覆盖的节点个数、平均出警时间、工作量平均值．由于先不考虑管辖区域的分配问题，此时工作量的方差不作为这里的评价指标．

而交巡警平台设置不合理又分有两种情况：一是平台数量不够，二是平台位置不合理．对于前者，主要的解决方法是新增交巡警平台设置；对于后者，主要的解决方法是调整现有的交巡警平台位置．当然，也有可能两种不合理的情况同时存在，但从节省人力资源的角度考虑，在能够调整交巡警平台位置的情况下，尽量减少新增平台数量．

1) 最短路算法下对时间的分析，见表4所示．

2) 对平均工作量的分析，见表5所示．

综上两种分析表格，以A区数据为参考指标，可以初步做出判断，如表6所示．

表4: 各地区出警情况统计

	最短路算法下的时间量的分析					
城 区	A	B	C	D	E	F
最长出警时间 (min)	57	44.7	68.6	160.6	191	84.8
无法3分钟到达节点个数	6	6	47	12	33	35
平均出警时间 (min)	11.22	15.27	23.35	25.06	25.38	24.3

表5: 各地区平均工作量统计

	平均工作量					
城 区	A	B	C	D	E	F
平均工作量	6.225	8.3	11.01	7.53	7.96	9.93

表6: 各地区解决方案

区域	主要问题	解决方案
B区	情况与A区相类似	增加、调整交巡警平台方案结合
C, F区	交巡警平台工作量过大，在现有警力资源条件下，交巡警平台的负担过重	以增加新的交巡警平台为主
D, E区	交巡警平台位置设置得不合理，因此产生部分节点无法被覆盖的情况	以调整现有交巡警平台位置为主

3.4.2 各城区评价及改进模型的建立

以A区为标准参照模板，对于各区的评价标准可以参考A区地方案进行．

建立一个新增平台的评价模型，当新增平台数为i时，分别用x_i, σ_i^2, t_i表示工作量期望、工作量方差和平均出警时间；另外定义x_0, σ_0^2, t_0为初始值，即在新增平台数为0的情况下的工作量期望、工作量方差及平均出警时间．

x_i衡量工作量压力，σ_i^2衡量工作量分配的均衡度，t_i衡量出警效率，将这三者加权值可以综合评定出模型的优劣．因此我们定义一个评价因子

$$F=\lambda_1\frac{x_i}{x_0}+\lambda_2\frac{\sigma_i^2}{\sigma_0^2}+\lambda_3\frac{t_i}{t_0},$$

其中$\lambda_1+\lambda_2+\lambda_3=1$．目标为：$\min F$, $\lambda_1,\lambda_2,\lambda_3$的取值根据实际情况和具体要求确定(例如，若更注重工作量的均衡度，可取$\lambda_1=\frac{1}{4}$, $\lambda_2=\frac{1}{2}$, $\lambda_3=\frac{1}{4}$)．

上述评价模型中没有涉及到最长出警时间的参数，因为分析表4的数据可知，最长出警时间与平均出警时间之间具有一定的相关性，而前者是特殊情况，后者则是基于整体的一个考虑，因此作为评价指标更有意义．

3.4.3 具体分析各个地区交巡警平台分布

先增加平台．采用前面平台增加的模型，在未设置平台的节点总数$N+n-M$情况下，对新增的m个平台的位置进行枚举，并且对每种方案计算出对应的评价因子F值，选取F值最小的方案．

以B区为例，取$\lambda_1=\lambda_2=\lambda_3=\frac{1}{3}$，计算出增加不同平台数时巡警的各项指标，并根据各项指标数据以及F值的计算公式，可画出F值随平台增加个数变化的曲线图，如图5所示．

根据走势图，再结合计算数据，当增加平台数大于等于三个时，没有出警时间大于三分钟的节点，因此选择新增三个平台为最佳，这三个平台的位置是123, 150, 159.

同理，可对C, D, E, F区作类似处理，得到各城区应增加平台的节点统计表，如表7所示．绘制出新增平台后的平台设置图，如图6所示．

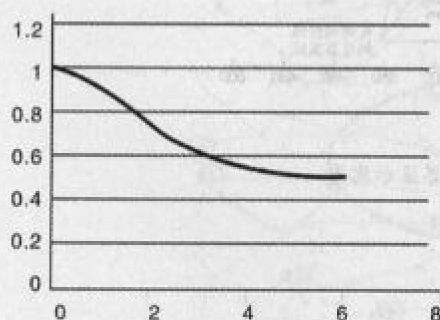


图5: B区评价因子F走势图

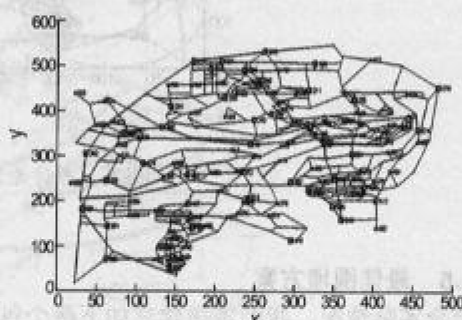


图6: 全城范围内新增平台后的平台设置图

表7: 各城区应增加平台的节点统计表

城区	增加平台的节点	增加平台个数
A	28 36 48 57 81	5
B	123 150 159	3
C	195 317 314 242 240 256 261 301 288 280 275 203 206 211	14
D	331 339 357 362 369	5
E	388 391 407 474 413 452 418	7
F	487 495 508 510 548 524 528 512 561 574 578	11

3.4.4 综合分析各城区交界区域的交巡警平台设置

分区考虑时，认为每个城区的警力只负责自己城区内的节点，而在实际情况中，警力是受到整个城市指挥系统的调配的，对交巡警平台警力进行跨城区抽调时，采取"就近原则"．因此我们对各个城区交界处的警力分布进行进一步调配，主要考虑城区交界处警力的合并．

第1步 对全市交界处的平台和节点进行搜索，找出交界区域；

第2步 将交界处看做是一个区，对其采用划分模型的做法，然后计算出增加的平台个数及其位置；

第3步 对交界区域的交巡警平台设置进行优化.

利用3.1和3.3的评价模型，对交界区域的交巡警平台分布进行评价，如图7所示. 得出结果是不需要新增平台，因此采取撤销两区部分交巡警平台的策略，对交界区域的平台进行优化. 撤销的交巡警平台编号为48, 100, 372.

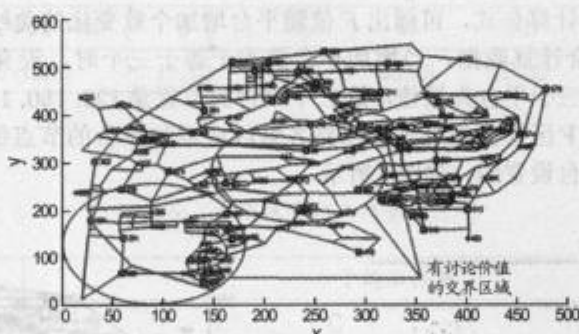


图7: 有讨论价值的交界区示意图

3.5 最佳围堵方案

结合实际情况，我们考虑建立如下两个包围圈：

城市包围圈: 即封锁圈. 在接到报案后对全城进行封锁.

内包围圈: 即追捕圈. 计算罪犯从案发地中心开始发散可能逃离的范围. 定一个小的包围圈去将罪犯封锁住，尽量将罪犯控制在较小的范围内.

外包围圈的封锁: 根据封锁模型，建立外包围圈，经计算得，对全市进行封锁的时间为12.364min. 如图8所示.

内包围圈的封锁: 为减少罪犯的危害，提高工作的效率，应尽量将罪犯限制在小范围内，尽快地将其抓住.

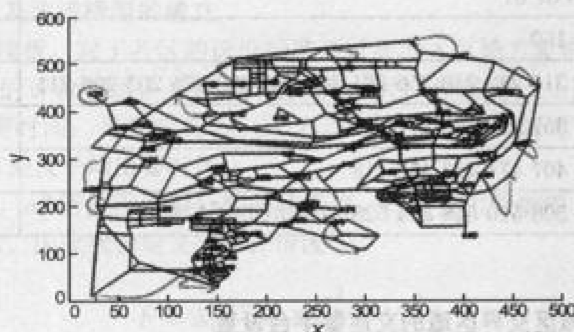


图 8: 市区外围的封锁方案

小范围封锁的具体算法如下：

初始化$t=0$，其中t为从报案开始后经过的时间；

步骤1：$t = t + 0.1$，计算出需要封锁的点的集合S；

步骤2：对于S中的点，找在时间范围内分别可以有哪些巡警站点能覆盖到；

步骤3：用匈牙利算法判断是否为X完全匹配，是则成功，实现了封锁；否则返回步骤1.

对于所给题目，$t' = 3\,\text{min}$. 再从犯罪人的心理考虑，其会在尽量避免被交巡警发现的前提下尽量提高车速，经查阅资料，得知各城市限速基本为60km/h. 所以假设逃跑犯罪嫌疑人的逃跑速度为60km/h. 假设警车的速度也是60km/h. 然后用计算机进行模拟，可得出经过6.2min时，可将罪犯封锁住. 具体封锁方案如图9所示，其中图中红色箭头$a \to b$表示交巡警从a出发到b点去封锁路线；红色的圈表示在封锁区域；蓝色的边表示割边，红色的边表示可以完全覆盖的点.

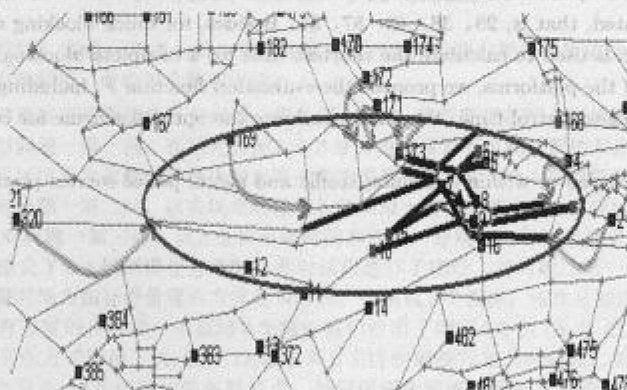


图9：小范围包围图

参考文献：

[1] 秦裕瑗，秦明复. 运筹学简明教程[M]. 北京：高等教育出版社，2006

Qin Y Y, Qin M F. A Short Course on Operations Research[M]. Beijing: Higher Education Press, 2006

[2] 张莹. 运筹学基础[M]. 北京：清华大学出版社，2010

Zhang Y. The Basis of Operations Research[M]. Beijing: Tsinghua University Press, 2010

[3] 殷剑宏，吴开亚. 图论及其算法[M]. 合肥：中国科学技术大学出版社，2003

Yin J H, Wu K Y. Graph Theories and its Algorithms[M]. Hefei: University of Science and Technology of China Press, 2003

[4] 卜月华. 图论及其应用[M]. 南京：东南大学出版社，2000

Bu Y H. Graph Theories and its Applications[M]. Nanjing: Southeast University Press, 2000

[5] 胡伯涛. 最小割模型在信息学竞赛中的应用[M]. 北京：NOI冬令营论文集，2007

Hu B T. Applications of Minimum Cut Model in Informatics[M]. Beijing: NOI Winnter Camp, 2007

Setting and Scheduling of Traffic and Patrol Police Service Platform

XING Zhong-yang, ZENG Fan-zheng, ZHANG Jiang-bin

(College of Opto-Electronic Science and Engineering, National University of Defense Technology, Changsha 410073)

Abstract: In this paper, we investigate the scheduling and setting of the traffic and patrol police service platforms, aiming to optimize the patrol time and balance the workload for each platform. The Floyd algorithm is used to attain the shortest route, and then the workload of each platform is optimally compromised. Accordingly, each of their optimal administration area can be adaptively allocated and the location of added platforms can be generated, that is, 28、36、48、57、81. Besides, for quick blocking of the 13 traffic arteries, the Hungarian algorithm is used to calculate the shortest time for a complete blocking, namely, 8.015 minutes. As for the adjustment of the platforms, we propose the evaluation function F, including the expected workload, the variance, and the longest patrol time. We finally evaluate the optimal scheme for besieging suspects within the blockade circle.

Keywords: shortest path; cover within 3 minute; traffic and patrol police service platform

第28卷 增刊1
2011年11月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 28 Supp.1
Nov. 2011

文章编号:1005-3085(2011)07-0145-08

设置与调度交巡警服务平台的模型建立

王 伟, 李 新, 章建鹏

指导教师: 罗文昌

(宁波大学理学院, 宁波 315211)

编者按: 该文的建模非常简洁、规范、考虑全面, 并作各种比较. 在评价现有平台设置的合理性时, 还建立了对所有路口重新设置平台的模型, 并给以求解和评价.

摘 要: 本文考虑了设置与调度交巡警服务平台问题. 首先用Matlab编程获得各节点间最短路矩阵, 然后将其转化为最短时间矩阵. 对问题一第一问, 在预先分配6个3分钟内交巡警不可能赶到的节点后, 以3分钟出警时间为约束, 分别建立了以出警次数最多的最小和出警次数方差最小为目标的0-1规划模型, 通过比较得出后者的出警次数更均衡. 对问题一第二问, 以实现全封锁最长时间最小为目标建立了0-1规划模型, 求得最长封锁时间为8.0155分钟. 对问题一第三问, 以3分钟出警时间为约束, 分别以出警次数最多的最小和出警次数方差最小为目标函数, 建立了0-1规划模型并求解, 并对结果进行了比较. 对问题二第一问, 以出警时间, 出警次数均衡性和总出警时间为指标评价现有方案是否合理. 根据这3个指标, 依次分别建立了3个优化后的0-1规划模型来评价现有方案的合理性, 依据这3个指标我们给出了相应6个区 (A, B, C, D, E, F) 的评价结果, 用重设平台或增设平台方式给出了A, B, C, D, E, F共6个区的调整方案. 对问题二第二问, 先用Matlab编程找出从P (第32个节点) 出发可到达的各层节点, 然后用逐步退缩围堵的思想给出了围堵方案, 此方案最多需7.362分钟即可围堵住犯罪嫌疑人.

关键词: 平台设置和调度; 0-1规划; 出警时间; 出警次数

分类号: AMS(2000) 90C30 中图分类号: O221.2 文献标识码: A

1 基本假设

1) 车辆行驶畅通无阻, 警车途中不发生拥堵等异常情况;
2) 在每个路口节点配置的警力及设置的交巡警服务平台的费用相当, 差别不大;
3) 交巡警按最短路径驾驶车辆前往案发路口节点;
4) 犯罪嫌疑人行驶速度为60 km/h;
5) 路口节点间的道路为双向道路.

2 问题一模型的建立与求解

2.1 建模准备: 构建最短时间矩阵

利用附件中所给节点坐标及线路数据, 用Matlab编程处理, 并利用Floyd算法[1]求出所给交通网络图中任意两点间的最短路矩阵 $(d_{ij})_{n\times n}$. 然后将最短路矩阵转化为最短时间矩阵 $(t_{ij})_{n\times n}$.

2.2 第一问的解答

首先提取A区中交巡警服务平台到路口节点的最短时间矩阵 $(t_{ij})_{20\times 92}$, 通过分析矩阵 $(t_{ij})_{20\times 92}$ 可发现, 在现有的20个交巡警服务平台设置下, 交巡警3分钟内不能到达的路口节点标号为28, 29, 38, 39, 61和92. 因此, 先把这些路口节点预先分配给距离它们最近的

交巡警服务平台，然后以3分钟出警时间为限，以出警次数最多的最小或出警次数方差最小建立模型2.1和2.2如下[2].

模型2.1

$$\min\ k \tag{1}$$

$$\text{s.t.}\ \sum_{j=1}^{92} x_{ij}p_j \le k,\quad i=1,2,\cdots,20, \tag{2}$$

$$x_{ij}t_{ij} \le 3,\quad i=1,2,\cdots,20,\quad j\in\{1,2,\cdots,92\}\backslash\{28,29,38,39,61,92\}, \tag{3}$$

$$\sum_{i=1}^{20} x_{ij}=1,\quad j=1,2,\cdots,92, \tag{4}$$

$$x_{15,28}=x_{15,29}=x_{16,38}=x_{2,39}=x_{7,61}=x_{20,92}=1, \tag{5}$$

$$x_{ij}=1,\quad 若\ i=j,i,\quad j=1,2,\cdots,20, \tag{6}$$

$$x_{ij}\ 为0或1,\quad i=1,2,\cdots,20,\quad j=1,2,\cdots,92, \tag{7}$$

其中i表示A区交巡警服务平台的位置标号；j表示A区交通网络中路口节点的标号；x_{ij}为0-1决策变量，若第i个交巡警服务平台服务第j个路口节点则为1，否则为0；k表示平台中最大出警次数；p_j表示第j个路口节点的案发次数. (1)式为模型的目标函数，其含义是极小化平台中最大的出警次数，该目标主要是考虑出警次数的均衡性；(2)式表示每个平台的出警次数不超过k；(3)式表示交巡警要在3分钟内从其所在平台到达其管辖的路口节点(除预先分配的路口节点外)；(4)式表示每个路口节点都有平台管辖；(5)式表示预先分配的路口节点被距离各自最近的服务平台所管辖；(6)式表示位于交巡警服务平台上的路口节点都由该服务平台自己来管辖.

利用Lingo软件[3]求解模型2.1可得所需结果，具体路口节点分配方案略.

模型2.2

$$\min\ \sum_{i=1}^{20}\Big(\sum_{j=1}^{92} x_{ij}p_j-\bar{p}\Big)^2$$

$$\bar{p}=\sum_{j=1}^{92} p_j\,/\,20,$$

其他约束条件同模型2.1中的(3)–(7)式,

其中$\bar{p}$表示出警次数的均值.

同样用Lingo软件求解模型2.2可得所需结果，具体路口节点分配方案略.

比较模型2.1和模型2.2的结果可知(具体细节略去)，模型2.2求得的方案出警次数均衡性更好.

2.3 第二问的解答

根据最短时间矩阵，提取出20个交巡警服务平台到出入A区的13个路口节点的最短时间矩阵$(t_{ij})_{20\times13}$，下面建立模型2.3.

模型 2.3

$$\min T \tag{8}$$

$$\text{s.t.} \sum_{i=1}^{20} x_{ij} = 1, \quad j = 1, 2, \cdots, 13, \tag{9}$$

$$x_{ij} t_{ij} \le T, \quad i = 1, 2, \cdots, 20, \quad j = 1, 2, \cdots, 13, \tag{10}$$

$$\sum_{j=1}^{13} x_{ij} \le 1, \quad i = 1, 2, \cdots, 20, \tag{11}$$

$$x_{ij} \text{ 为0或1}, \quad i = 1, 2, \cdots, 20, \quad j = 1, 2, \cdots, 13, \tag{12}$$

其中T为实现全封锁所需最长时间．

用Lingo求解可得实现全封锁所需最长时间为8.015457分钟，具体封锁方案略．

注1　对模型2.3我们可进一步改进使得实现全封锁所用的总出警时间最小．用模型2.3得到的T值作为约束，以实现全封锁所用的总出警时间最小为目标再建立一个0-1规划模型进行求解即可．

2.4　第三问的解答

在给A区增加服务平台时，应满足出警时间短和出警次数均衡这两个原则．出警时间短要求交巡警在3分钟甚至更短的时间内赶到案发现场，出警次数均衡原则要求每个平台出警次数尽量均衡．据此建立模型2.4及模型2.5如下：

模型 2.4

$$\min k \tag{13}$$

$$\text{s.t.} \sum_{j=1}^{92} x_{ij} p_j \le k, \quad i = 1, 2, \cdots, 20, \tag{14}$$

$$2 \le \sum_{i=21}^{92} m_i \le Z, \quad Z = 2, 3, 4, 5, \tag{15}$$

$$\sum_{i=1}^{92} x_{ij} = 1, \quad j = 1, 2, \cdots, 92, \tag{16}$$

$$x_{ij} t_{ij} \le 3, \quad i = 1, 2, \cdots, 20, \quad j = 1, 2, \cdots, 92, \tag{17}$$

$$x_{ij} \le m_i, \quad i = 1, 2, \cdots, 20, \quad j = 1, 2, \cdots, 92, \tag{18}$$

$$m_i = 1, \quad i = 1, 2, \cdots, 20, \tag{19}$$

$$x_{ij} = 1, \quad \text{若 } i = j, \quad i, j = 1, 2, \cdots, 20, \tag{20}$$

$$x_{ij} \text{ 为0或1}, \quad i = 1, 2, \cdots, 20, \quad j = 1, 2, \cdots, 92, \tag{21}$$

$$m_i \text{ 为0或1}, \quad i = 21, 22, \cdots, 92, \tag{22}$$

其中 m_i 表示是否设置第 i 个路口节点为交巡警服务平台，若是则为1，否则为0；Z 表示增加平台数量的上限值．目标为极小化平台中最大出警次数．(15) 式表示从标号为21至92的路口节点中选出2至5个设为服务平台．

模型 2.5

$$\min \sum_{i=1}^{92} m_i\Big(\sum_{j=1}^{92} x_{ij}p_j - \bar{p}\Big)^2 \tag{23}$$

$$\bar{p} = \sum_{j=1}^{92} p_j \Big/ \sum_{i=1}^{92} m_i, \tag{24}$$

其他约束条件同模型2.4中的 (15)–(22) 式，　　(25)

其中 $\bar{p}$ 表示出警次数均值，目标为极小化出警次数方差．

用Lingo软件求解模型2.4，将 Z 设定为2, 3时，无解．将 Z 设定为4, 5时，有解．将 Z 设定为5，新增平台应设在标号为21, 29, 40, 48和88的路口节点；将 Z 设定为4，新增平台应设在标号为29, 39, 48和88的路口节点．通过比较 (具体细节略去)[4]，我们发现多增设一个平台对出警次数均衡性并没有明显的改善，多增设的那个平台只是分担了那些能正常出警的平台的工作，效果不好，反而增加了财政负担．在这种情况下，一般新增4个平台比较合适．当然在财政充裕的情况下增设5个平台也是可以的．

注 2　对模型2.4，我们可进一步改进使得所用的总出警时间最小．用模型2.4得到的 k 值作为约束，以总出警时间最小为目标建立一个0-1规划模型进行求解即可．类似的分析适用于模型2.5．

3　问题二模型的建立与求解

3.1　第一问的解答

1)　评价指标体系的建立

针对全市的具体情况，为了准确地评价交巡警服务平台设置方案的合理性，首先得建立一套正确的评价体系．从交巡警服务平台的原则和任务出发，本文选用如下3个指标：

指标1：　出警时间

出警时间是指从交巡警服务平台到达其所管辖的路口节点所用的时间，该指标越小越好．

指标2：　出警次数均衡性

该指标是使各交巡警服务平台的出警次数尽量均衡，可以用极小化平台出警次数的方差或平台中出警次数最多的最小表示，即 $\min\{\max_i \sum_j x_{ij}p_j\}$．

指标3：　总出警时间

总出警时间是指所有交巡警服务平台赶往案发点所花时间之和，即 $\sum_i \sum_j x_{ij}t_{ij}p_j$．

在实际中，出警时间最重要，出警次数均衡性次之，总出警时间作用最弱．

2)　合理性评价

为评价现有方案，我们先对交巡警服务平台采用重设平台的方式进行优化，通过比较现有方案和优化后方案的指标，确定现有方案是否合理，是否需要改进．为此建立优化模型如下：

模型 3.1　根据指标1建立最长出警时间最短模型

模型如下所示

$$\min T \tag{26}$$

$$\text{s.t.} \sum_{i=1}^{a_z} x_{ij} = 1, \quad j = 1,2,\cdots,a_z, \tag{27}$$

$$x_{ij}t_{ij} \le T, \quad i = 1,2,\cdots,a_z, \quad j = 1,2,\cdots,a_z, \tag{28}$$

$$x_{ij} \le m_i, \quad i = 1,2,\cdots,a_z, \quad j = 1,2,\cdots,a_z, \tag{29}$$

$$\sum_{i=1}^{a_z} m_i = b_z, \tag{30}$$

$$x_{ij} \text{ 为0或1}, \quad i = 1,2,\cdots,a_z, \quad j = 1,2,\cdots,a_z, \tag{31}$$

$$m_i \text{ 为0或1}, \quad i = 1,2,\cdots,a_z, \tag{32}$$

其中 $z = 1,2,\cdots;6$ 分别表示该市 A, B, C, D, E, F 六个区；b_z 表示第 z 区现有的平台数量，a_z 表示第 z 区的路口节点数．约束条件 (30) 表示新设置的平台总数和现有方案相等．

模型 3.2　根据指标2建立出警次数均衡性模型

以平台中出警次数最多的最小为目标，并把模型 3.1 中得到的 T 值作为约束，建立模型如下

$$\min k$$

$$\text{s.t.} \sum_{j=1}^{a_z} x_{ij}p_j \le k, \quad i = 1,2,\cdots,a_z,$$

其他约束条件同模型 3.1 中的 (27)–(32) 式，

或以平台出警次数的方差为目标，用模型 3.1 得到的 T 值作为约束，建立模型如下

$$\min \sum_{i=1}^{a_z} m_i\Big(\sum_{j=1}^{a_z} x_{ij}p_j - \bar{p}_z\Big)^2$$

$$\text{s.t.} \quad \bar{p}_z = \sum_{j=1}^{a_z} p_j \Big/ \sum_{i=1}^{a_z} m_i,$$

其他约束条件同模型 3.1 中的 (27)–(32) 式．

模型 3.3　根据指标3建立总出警时间最少模型

以总出警时间最少为目标，并把模型 3.1 和模型 3.2 得到的 T 值和 k 值作为约束，建立模型

如下

$$\min \sum_{i=1}^{a_z}\sum_{j=1}^{a_z} x_{ij}t_{ij}p_j$$

$$\text{s.t.} \quad \sum_{j=1}^{a_z} x_{ij}p_j \le k, \quad i=1,2,\cdots,a_z,$$

其他约束条件同模型3.1中的(27)–(32)式.

用Lingo软件求解可得所需结果，见表1及表2．比较表1和表2，发现用重设平台方式得到的优化方案在出警时间方面有了很大的改善，而其他的指标变化不是很明显．这说明现有方案不够合理需要改进．在实际中，出警时间起着决定性的作用，这是设立交巡警服务平台的第一原则．只有在满足该原则下，考虑出警次数均衡性才有意义．

表1: 现有方案的各个指标

区号	A	B	C	D	E	F
出警时间	5.7005	4.470312	6.8605	16.06282	19.1051	8.47985
最多出警次数	7	8.4	13.1	7.7	8.4	10.8
出警次数方差	1.8919	0.0175	10.8905	0.1133	0.5557	2.6165
总出警时间	142.90	97.01	403.307	176.3078	379.44	293.16

表2: 以重设平台方式得到的优化后方案的各个指标

区号	A	B	C	D	E	F
出警时间	3.4713	3	4.6	6.8529	5.0221	7.5932
最多出警次数	7	8.4	13.1	8.5	9.5	10.8
出警次数方差	1.9949	0.0275	8.4634	4.8911	7.7637	3.5147
总出警时间	135.6697	73.97	344.54	121.55	306.21	223.32

3) 优化后调整方案

观察表2可知，A区和B区出警时间基本上在3分钟之内，增设平台又需要成本，所以本文认为A区和B区通过重设平台就能得到较合理的方案．对于C, D, E, F区则需要用增设平台方式来优化，这只要在模型3.1、模型3.2和模型3.3中逐步扩大b_z的值即可．得到的评价结果见表3．比较表2和表3，发现用增设平台的方式得到的优化方案在出警时间上有了较大的改善，而其他的指标也有一定程度的改善．具体调整方案见表4.

在财政紧张的情况下，用重设平台方式也能得到较好的结果，其调整方案见表5.

3.2 第二问的解答：围堵方案(略)

表3: 以增设平台方式得到的优化后方案的各个指标

区号	C	D	E	F
出警时间	4	4.6057	4.2202	4.1012
最多出警次数	8.4	7.5	8	8
出警次数方差	6.4316	6.7758	2.9893	6.0333
总出警时间	274.89	130.44	247.53	176.26

表4: 以增设平台方式得到的优化后方案

区号	各交巡警服务平台位置所在节点标号
C	166, 167, 171, 175, 179, 186, 191, 201, 210, 219, 224, 225, 227, 244, 247, 254, 261, 269, 271, 276, 287, 291, 296, 311, 316
D	321, 324, 325, 330, 334, 337, 343, 345, 347, 361, 362, 370
E	373, 375, 379, 385, 387, 388, 391, 398, 403, 408, 419, 441, 446, 453, 457, 458, 466, 468
F	475, 478, 479, 482, 483, 485, 492, 507, 521, 525, 540, 548, 553, 372, 567, 575, 577, 582

表5: 以重设平台方式得到的优化后方案

区号	各交巡警服务平台位置所在节点标号
A	3, 4, 5, 9, 10, 17, 18, 20, 21, 24, 27, 29, 31, 36, 40, 48, 59, 69, 76, 89
B	99, 102, 106, 116, 123, 133, 147, 164
C	167, 169, 170, 173, 175, 179, 186, 201, 210, 213, 228, 240, 262, 272, 286, 307, 314
D	320, 324, 326, 329, 334, 341, 361, 367, 170
E	384, 387, 388, 391, 396, 403, 405, 418, 423, 429, 448, 450, 456, 457, 460
F	489, 498, 511, 518, 539, 542, 548, 553, 557, 569, 576

参考文献：

[1] 谢金星，邢文训，王振波. 网络优化(第2版)[M]. 北京：清华大学出版社，2009
Xie J X, Xing W X, Wang Z B. Network Optimization (2nd Edition)[M]. Beijing: Tsinghua University Press, 2009

[2] 姜启源，谢金星，叶俊. 数学模型(第3版)[M]. 北京：高等教育出版社，2003
Jiang Q Y, Xie J X, Ye J. Mathematical Model (3rd Edition)[M]. Beijing: Higher Education Press, 2003

[3] 袁新生，邵大宏，郁时炼. Lingo和Excel在数学建模中的应用[M]. 北京：科学出版社，2007
Yuan X S, Shao D H, Yu S L. Application of Lingo and Excel in Mathematical Modeling[M]. Beijing: Science Press, 2007

[4] 薛定宇，陈阳泉. 高等应用数学问题的Matlab求解(第2版)[M]. 北京：清华大学出版社，2008
Xue D Y, Chen Y Q. Matlab Algorithm of Higher Applied Mathematics Problems (2nd Edition)[M]. Beijing: Higher Education Press, 2008

Modeling the Setting and Scheduling of Traffic and Patrol Police Service Platforms

WANG Wei, LI Xin, ZHANG Jian-peng
Advisor: LUO Wen-chang

(Faculty of Science, Ningbo University, Ningbo 315211)

Abstract: In this paper, we consider the problem of setting and scheduling of traffic and patrol police service platforms. First, we obtain the shortest path matrix of nodes by programming with the Matlab software, and then transform it into the shortest time matrix of nodes. For the first question of Problem one, we pre-assign 6 nodes that the police cannot arrive at within 3 minutes and then propose the corresponding 0-1 programming models with the goals to minimize the maximal workload of police and the variance of workload of police under the constraint that the responding time of police is no more than 3 minutes, respectively. By comparing these results, we show that the latter is better on compromising the workload of police. For the second question of Problem one, we present a 0-1 programming model with the goal to minimize the longest responding time of police subject to blockading all the 13 access arteries. By solving the proposed model, we show that the longest blockading time is 8.0155 minutes. For the third question of Problem one, we derive the corresponding 0-1 programming models with the goals to minimize the maximal workload of police and the variance of workload of police under the constraint that the responding time of police is no more than 3 minutes. Then the results are compared. For the first question of Problem two, we propose three indexes, i.e., the responding time of police, the workload of police and the total responding time of police, to evaluate the reasonableness of the present schemes, respectively. According to the three indexes, we propose the corresponding 0-1 programming models step by step to evaluate the reasonableness of present schemes. Based on the three indexes, we give the evaluation results for the six districts (A, B, C, D, E, F). By resetting the platforms or adding the new platforms, we give the adjustment schemes for the six districts (A, B, C, D, E, F). For the second question of Problem two, we first get the nodes of each level from P node to arrive at by programming with Matlab. Then by using the idea of gradually giving back to contain the suspects, we propose the containment strategy. The suspects can be contained within 7.362 minutes by using the new strategy.

Keywords: setting and scheduling of platforms; 0-1 programming; responding time of police; workload of police

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0153-06

交巡警服务平台的设置与调度

张 潇， 田志祥， 傅建龙

指导教师： 教练组

(上海财经大学金融学院，上海 200433)

编者按：论文在抓捕犯罪嫌疑人时的分析、思路和算法都十分清楚，并考虑了嫌疑人逃跑时不同速度对围堵范围的影响，还对围堵成功后的搜捕过程进行了讨论．在封锁方案的制订部分和犯罪嫌疑人围堵过程中的平台调度都用最大流的方法进行计算，显得处理方法统一．故我们选择将答卷的这部分内容发表，供大家参考．

摘 要：本文首先利用Floyd算法对交巡警平台进行初步分配，并利用网络流模型制定出封锁方案．其次针对警务资源的优化问题设立综合评价模型，对目前的分配状况进行评价并提出改进．最后分步建立了围堵与抓捕模型，使用网络流的方法求解，得到嫌犯不同最大车速下的不同围堵方案，再采用逐步缩小包围圈的方法，在邻点度约束和连通性约束的条件下采用基于准备时间的贪心调整法，在一小时内抓获嫌犯．

关键词：网络流模型；双目标规划模型；模拟退火算法；综合评价模型

分类号：AMS(2000) 90B10 中图分类号：O22 文献标识码：A

1 问题重述 (略)

2 模型假设

1) 突发事故都发生在路口节点上；
2) 每个路口只被一个交巡警服务平台管辖；
3) 相邻两个交叉路口之间的道路近似认为是直线；
4) 所有道路都是双向且畅通无阻的，警车由最短路径到达事发地点；
5) 警车以60 km/h的速度匀速行驶，不考虑车辆的调头、启动、停止时的加减速过程．

3 符号说明

V：图中所有顶点的集合；

E：图中所有路线(边)的集合；

$G(V,E)$：抽象的整张无向图；

$\min(i,j)$：从节点i到节点j的最短距离(时间)；

$t_{\max}$：最长出警时间$[=\sup\min(i,j)]$；

r_i：平台i管辖区域内的总发案率．

4 问题一的模型

4.1 管辖范围的分配

此问要求路口节点能够尽快从所属的平台到达，在不考虑均衡各平台工作量的情况下(将在下文讨论)，只需将各路口分配给离它最近的平台管辖，利用Floyd算法为A区20个交巡警平台合理分配了管辖范围.

4.2 封锁方案的制定

此问要求调度20个平台的警力来封锁A区的13个出口，每个平台的警力只能封锁一个出口，而每个出口也只需要一个平台．由于这种一一对应的关系，很自然地将问题化归为典型的匹配模型．将每个封锁方案看作一个向量$(x_1, x_2, \cdots, x_{13})$，$x_i \neq x_j$．平台$x_i$用来封锁出口$y_j$．优化指标为：$\min\max m(x_i, y_j)$，$i = 1, 2, \cdots, 13$.

像这样定量的匹配模型难以直接求得最优解，因而需要对问题作进一步转化．考虑先将问题转化为判定性问题，即对于一定的时间限制，能否找到一种匹配方案成功封锁；然后再考虑缩小时间上限，逼近最优解．设定了时间阈值后，网络流模型能很好地解决判断是否能在此时间内封锁所有路口这一问题．网络流模型的流量平衡条件能够保证每个平台只封锁一个路口，而最大流就是最多封锁的路口数．将此算法记为算法一，具体的执行步骤如下：

算法一：

步骤1 从源到所有平台连一条流量上限为1的边；

步骤2 从所有要封锁的出口向汇连一条流量上限为1的边；

步骤3 对于任意一对平台和出口，若它们之间的最短时间小于当前的阈值，则从平台向出口连一条流量上限为1的边；

步骤4 计算模型的最大流；

步骤5 若流量等于出口数，则所有出口在规定时间内均被封锁.

用二分法求得满足条件的最小时间阈值为8.0157 (m)．相对应的警力调度方案如表1.

表1: 快速封锁警力调度方案

路口标号	12	14	16	21	22	23	24
平台编号	A10	A16	A2	A11	A12	A14	A13
路口标号	28	29	30	38	48	62	
平台编号	A15	A7	A5	A1	A4	A3	

4.3 服务平台的增设

此问要求根据现有交巡警服务平台的工作量不均衡和有些地方出警时间过长的实际情况增设服务平台．建立以最长出警时间及工作量均衡度为目标的双目标规划模型并采用宽容分层序列法求解[1]．最终确定为A区增加4个平台，位置分别为28, 40, 48, 91号节点.

5 问题二的模型

5.1 全市平台设置的合理性

此问要求分城区考虑平台设置方案的合理性．建立综合评价模型，提出出警时间达标率、平均超出时间以及平均工作量这三个评价指标，进而用夹角余弦法确定指标的权重．经求解比较后，认为A, B两区的平台设置方案相对合理，其余四区的方案相对不合理，其中D, E两区主要是位置安排不合理，C, F两区主要是警力不足．针对不合理各区的特定原因，我们给出了相应的解决方案[2,3].

5.2 抓捕犯罪嫌疑人

关于嫌犯，我们首先假设其踪迹在案发后不可追踪，即3分钟内嫌犯可以跑到路程范围允许的一切地方，而警方对此没有具体信息．其次，假设嫌犯充分狡猾，如果有任意一条逃离市区的道路没有在合适的时间内被封锁，抓捕工作都会失败．最后，假设嫌犯的车速可快可慢，可掉头、转弯，但最大车速固定，记作v．

抓捕工作分两步完成：首先迅速建立包围圈围堵嫌犯，使嫌犯无法逃出该市．若无法建立这样一个包围圈，则抓捕工作失败；其次在包围圈内搜捕嫌犯，搜捕工作的基本前提是不能放跑嫌犯，目标为在尽量短的时间内抓获嫌犯．若效率相当，则使用警力的越少越好．

5.2.1 围堵模型

在寻找评判是否能围堵成功之前，首先介绍两个定义：若点j属于点集S，且其所有邻点均属于S，则称点j为集合S的内点；反之，则称其为S的边界点．将内点集记作V_1，边界点集记作V_2．$S = V_1 + V_2$, $V_1 \cap V_2 = \emptyset$，若点$j \in V_2 \subset S$，将j的所有邻点加入S称为对点j的一次增广，此时$j \in V_1$．如果我们能够找到一个点集S，包含嫌疑犯的出发点，且集合S的所有边界点警察都能赶在嫌疑犯之前到达，则围堵成功；否则，围堵失败．我们希望在最小的地理范围内进行围堵．设计的围堵模型如图1所示，具体步骤如下：

步骤1　将嫌犯在案发3分钟内可能到达过的顶点(在路上不算)集定义为初始集，记作S_0；

步骤2　将S_0的所有边界点一起增广，新得到的点集记作S_1．显然$S_0 \subset S_1$，且S_0中所有的点均为S_1的内点，S_1是满足成功围堵条件的最小集合．嫌犯若要逃离，必须经过S_1的某个边界点，$V_2 \subset S_1$便形成了一个包围圈．因此，警方只要赶在嫌犯之前封锁住这些边界点，则围堵成功；

步骤3　若不能在目前的点集围堵成功，则选择其中一个边界点进行增广，再进行评判．在边界点的选择上，我们采用贪婪法，选择嫌疑犯到达时间和最近平台到达时间差最大的边界点，基本想法是增广最难包围的边界点．

对于当前的点集，可以采用算法一进行评判是否可以包围，当$v = 1$时的包围圈如图2所示．具体步骤如下：

步骤1　从源到所有的平台连一条流量上限为1的边；

步骤2　从所有边界点向汇连一条流量上限为1的边；

步骤3　若平台i到边界点j的时间$m(i,j) < m(32,j)/v - 3$，则从平台i向边界点j连一条流量上限为1的边；

步骤4　计算在给定的包络集下模型的最大流；

步骤5　若流量等于边界点数目，则围堵成功．

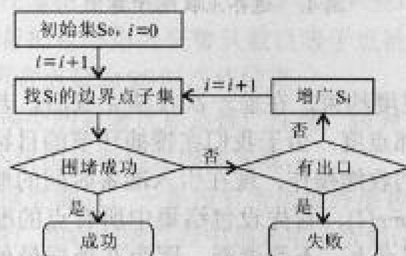


图1:　围堵模型

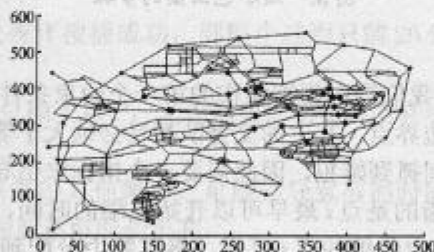


图2:　当$v = 1$时包围圈示意图

我们可以将围堵结果进行灵敏度分析，在参数v取不同值的情况下，我们得到不同的包围圈及其所含节点数，见表2．当疑犯的最大车速与警车速度相比较小时，用来围堵疑犯的包围圈所含节点数对速度的变化比较敏感；当疑犯车速与警车速度相当时，包围圈所含节点数表现出相当良好的稳定性；而当疑犯车速比警车速度快到一定程度时，我们发现在此问题中已经无法围堵疑犯．

表2: 包围圈所含节点数与嫌犯最大车速的关系

v (km/min)	0.5	0.6	0.7	0.8	0.9	1
节点数	9	12	20	22	20	21
v (km/min)	1.1	1.2	1.3	1.4	1.5	>1.5
节点数	22	21	21	22	22	不可

5.2.2 搜捕模型

将每个平台的警力称为一支交警队．围堵成功后，逐步缩小包围圈进行抓捕，初步考虑由所有参与围堵的交警队和包围圈内部平台的交警队联合进行．当警车与嫌犯车辆在道路上或路口节点相遇时，抓捕即宣告成功．若能每次成功减少包络集的一个边界点，且不放跑嫌犯，则在有限步操作后必能抓捕成功，具体流程图见图3．定义一个边界点邻接的内点数量为该边界点的内点度，定义一个边界点邻接的内点与边界点的数量总和为它的邻点度．首先考察最外层的包络集(即成功围堵的包络集)，所有内点度为1的边界点必能被收缩至与之邻接的唯一内点，即将该边界点的警力调度到邻接的唯一内点上．然而上述操作过后，再直接收缩内点度为1的边界点则有可能会放跑嫌疑人，原因是最外层包络集的边界点之间不可能存在嫌疑人，但是收缩之后的包络集的边界点之间有可能存在嫌疑人，如图4中右图．

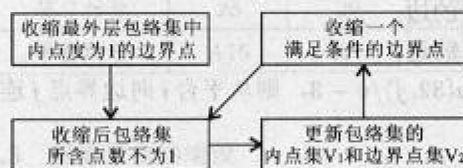


图3: 缩小包围圈的步骤

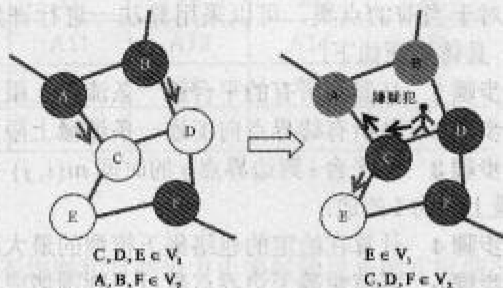


图4: 边界点收缩示意图

由此我们得出收缩操作的第一个约束条件：邻点度约束．在前一次收缩完成后，若要再收缩某个边界点，其上的交警队数目必须大于等于其邻点度．由于我们在搜捕环节的目标是花最少的时间抓到嫌犯，因此需要一个指标来引导我们的收缩操作．现在引入准备时间的概念．准备时间指的是点i最早可以开始收缩的时间，记作$pre(i)$．首先设包络集中所有点的准备时间为0．选定收缩的点i以后，将其警力分配到所有邻点上，不得遗漏，因为在处理最外层包络集后，嫌犯可能位于两个边界点之间．邻点j的准备时间更新为

$$pre(j)=\max\left\{pre(i)+t(i,j),pre(j)\right\}.$$

通过这个操作，虽然我们每次只收缩一个边界点，但实际上它们是可以同时进行的．每次只收缩一个点相当于给定一个收缩顺序，相当于每次交警队到达新的地点时就原地待命，直到下一步指示可以继续缩小包围圈．在收缩过程中互不影响的点同时进行，如图2所示的局部包围圈，边上的数字代表路程．A点的准备时间为5，假设它可以收缩，则B和C的准备时间更新为7和8，初始为0．紧接着收缩B点，D的准备时间更新为8．最后收缩C点，将D的准备时间更新为11．虽然每次只收缩了一个点，但是实际上模拟的是A的两批警力同时出发，并在D集合待命．

对于任意一种收缩顺序，如果最后其能顺利收缩为一个点，则最后一个点的准备时间即为搜捕时间，优化目标为 $\min pre(k)$，对于任意收缩顺序，其中 k 是最后的收缩点．

至此搜捕模型没有考虑没有可收缩边界点的情形．在编程求解的过程中，这个问题也确实发生了，当迭代到某一步，V_2 中没有任何一个点可以收缩，则搜捕失败．考虑到 G 是一张平面图，所以每个点的邻接点数都不太大，大多数小于5；相反，随着包围圈的缩小，警察越来越集中，收缩不应该在中途失败．这个矛盾的产生是因为在收缩的过程中，收缩某点有可能将包络集割裂．比如在图5中，若A和B收缩完毕后，D也成为了边界点，且可以收缩．当它收缩时一部分警力分配给了C，并将C孤立．C只能收缩至自身，其上含有的所有警力将无法继续参加包围，形成浪费．基于以上分析，我们得出收缩操作的第二个约束条件：连通性约束．每次收缩的边界点都不能破坏包络集的连通性．连通性是指集合内的任意两点，可以只通过集合内部的点互达．对于我们的优化目标 $\min pre(k)$，采用贪婪法选取收缩顺序，每次选择准备时间最小的可收缩边界点，将其警力分散到所有邻点，若警力不能被邻点数整除，则将余数任意分配至某邻点．

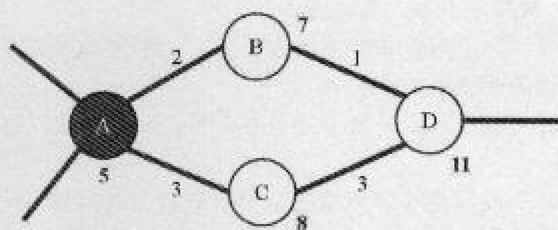


图5:　准备时间示意图

5.2.3　结果与优缺点分析

1)　在嫌犯车速为1时，抓捕过程共耗时56分钟．

2)　若把包络集内交警队数目等于点数的情况视作搜捕成功，则整个过程只需29分钟，当包络集所含点数为26时迭代结束．

3)　模型的显著优点是推广性强，因为它并不依赖于点P，无论嫌犯的初始位置在哪里，模型都能尽快围堵或判定围堵失败．

4)　模型缺点：一是没有考虑最外层包络集以外的其他警力，可能会导致搜捕时间偏长；二是在增广和收缩的过程中都采用了贪婪法，解的质量可能会不高．

5)　改进方向：在收缩或增广过程中，加入一些评估函数．但因为时间有限我们未能找到比贪婪法更好的智能搜索，实为遗憾．

参考文献:

[1] 张军. 计算智能[M]. 北京：清华大学出版社，2009

Zhang J. Computational Intelligence[M]. Beijing: Tsinghua University Press, 2009

[2] 吴礼斌. 数学实验与建模[M]. 北京：国防工业出版社，2007

Wu L B. Mathematics Experiment and Modeling[M]. Beijing: National Defence Industry Press, 2007

[3] 姜启源. 数学模型(第3版)[M]. 北京：高等教育出版社，2009

Jiang Q Y. Mathematic Model (3rd Edition)[M]. Beijing: Higher Education Press, 2009

The Optimal Setting and Scheduling of the Traffic and Patrol Police

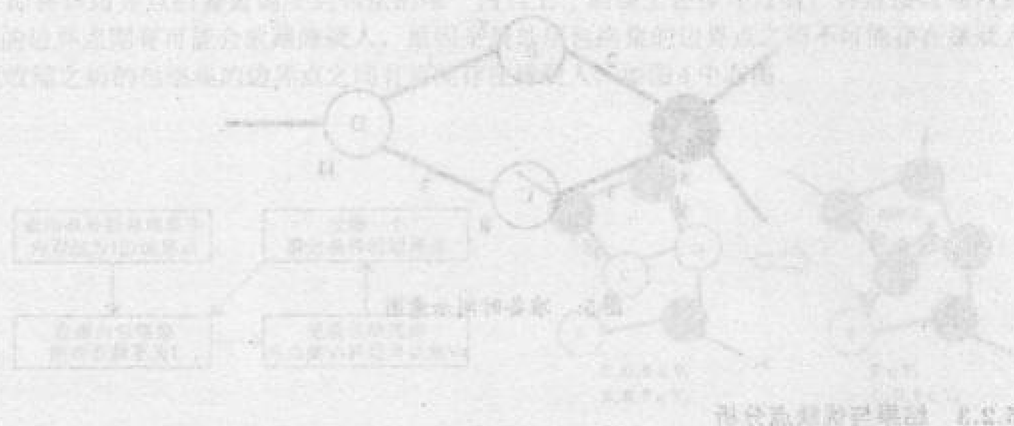
ZHANG Xiao, TIAN Zhi-xiang, FU Jian-long

Advisor: Instructor Group

(School of Finance, Shanghai University of Finance and Economics, Shanghai 200433)

Abstract: This paper first employs the Floyd algorithm to preliminarily allocate the police stations, and builds the network-flow model to establish a scheme to blockade an area. Second, a multi-objective programming model is used to evaluate the current allocation of the police resource and present the corresponding improvement. Finally an encircle-catch model is set up. The network-flow model is utilized again to make different encircling plans according to the suspect's speed, and on the basis of the greedy algorithm the encirclement under two constraints is constructed to catch the suspect within one hour.

Keywords: network flow model; multi-objective programming model; simulated annealing algorithm; comprehensive evaluation model



第28卷 增刊1
2011年11月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 28 Supp.1
Nov. 2011

文章编号:1005-3085(2011)07-0159-05

交巡警服务平台的设置与调度

刘 进, 杨 鋆, 吴 可
(武汉大学数学与统计学院，武汉 430072)

编者按: 本文在问题2.2(即第二大问第二小问的围堵方案)的处理上比较有特色，这里摘要发表. 论文首先给出迭代模型(每步迭代用整数规划建模)，然后给出迭代算法，最后用Matlab编程求解. 但算法设计和编程时没有严格按照模型求最优解，得到的应是启发式较优解. 值得一提的是：Matlab编程简洁、易懂，这里完整发表，有兴趣的读者不妨阅读并尝试运行之.

摘 要: 对于问题2.2，假设一个警力平台只能封堵一个节点，考虑到犯罪嫌疑人逃逸方向不确定，本文建立动态优化模型. 首先建立一个到事发点P距离为犯罪嫌疑人3分钟车程的节点集A，再以这个集上的点为出发点搜索出相邻点构造一个割集B，对割集B进行附近警力点的搜索和优化配置；若没有警力点满足封锁要求，则重新探索相邻点构造新的割集B，直到找到能够满足封堵的情况. 本文求出了犯罪嫌疑人的速度分别为60, 70, 80, 90 km/h的近似最佳围堵方案：当犯罪嫌疑人的速度为60 km/h，最短围捕时间为8.9 min.

关键词: 0-1规划；Floyd算法；平台调度；最佳围堵

分类号: AMS(2000) 90B10 中图分类号: O221.7 文献标识码: A

1 问题重述(略)

2 模型假设(略)

3 符号说明(略)

4 问题分析(略)

5 问题一模型的建立与求解(略)

6 问题二模型的建立与求解

6.1 问题2.1的模型建立与求解(略)

6.2 问题2.2的模型建立与求解

6.2.1 问题的分析与模型的建立

根据题意，本问要求最佳围堵方案，确定目标为调度全市交巡警服务平台警力资源使得围堵范围尽量小. 据题意，需要据现有交巡警服务平台设置方案进行求算，故本文在计算过程中使用原始数据. 简单起见不考虑在第一问中A区增设的新平台. 下设犯罪嫌疑人逃逸速度为v km/h，并记M_3为所有警力平台组成的集合，d_{jk}指节点j到节点k的距离. 模型建立如下：

1) 设M_1的初值为距离P点小于$v\cdot 3\,\mathrm{min}$(即$v/20$ km，下同)的路口节点组成的集合，即

$$M_1=\left\{i: d_{32,i}<v\cdot 3\,\mathrm{min}\right\}.$$

2) 设M_2为不在M_1中，且与M_1的节点直接相邻的路口节点组成的集合.

3) 建立以下0-1规划模型，寻求以M_2为围堵节点的最佳警力配置方案：

$$\min z = \max_{j,k}\{d_{jk}a_{jk}\}$$

$$\text{s.t.} \begin{cases} \sum\limits_{k} a_{jk} = 1, \\ \sum\limits_{j} a_{jk} \le 1, \\ d_{jk}a_{jk} \le d_{32,j} - v \cdot 3\,\text{min}, \\ a_{jk} = 0, 1, \quad j \in M_2, \quad k \in M_3, \end{cases}$$

其中，目标为最大围堵距离最短(等价于最大围堵时间最短)；决策变量a_{jk}指节点j是否由平台k去封堵；$\sum\limits_{j} a_{jk} \le 1$指平台$k$至多围堵一个节点；其他约束含义清楚，不再赘述.

若该问题有最优解，转下一步继续；否则(等价于没有可行解)，表明以M_2作为围堵节点时，任何警力配置下都不能保证将罪犯围堵，所以需要修改集合M_1，回到上一步继续.

寻找修改M_1的最佳方法，可能需要大量计算时间．可以近似计算警力配置方案(而不去求解上0-1规划)，并同时对M_1进行修改，例如简单地依次对$i \in M_2$，总是在M_3中找离节点i最近的平台k去封堵；若在i处不可以将犯罪嫌疑人围堵掉，即对任意未分配任务的平台k，$d_{ik} > d_{32,i} - v \cdot 3\,\text{min}$，则将$i$加入$M_1$中，回到上一步继续.

4) 如果上述优化问题有最优解，此时以a_{jk}为围堵方案，M_2中所有节点都能被围堵掉，则M_2为所要围堵节点，$M_1 + M_2$为围堵区域.

6.2.2 模型的求解

对上述模型在Matlab软件中求解，算法如下：

步骤1 以P为出发点，确定不能封锁的点的集合A的初值(3分钟之内的点)；

步骤2 确定与A相邻的点组成的集合B；

步骤3 考察B中的所有节点，如果B中任一节点不能及时封堵，则将该节点加入A，并跳转至第二步；

步骤4 如果B中所有节点都能在罪犯逃离该节点前被及时封堵(本文在编程中采用警力配置就近原则进行封堵)，则B即为所求封锁节点的集合，A+B为所求封锁区域.

分别取v =60, 70, 80, 90 km/h，可求得结果，如表1所示．当v = 60 km/h时，通过设防节点和围捕平台的相关标记得出围捕时间为8.9 min.

表1：不同逃逸速度下的围捕方案

60	设防节点	3 4 5 6 10 15 16 41 40 63 60 171 173 168 240 230 248 273 370 371 561
	围捕平台	3 4 5 6 10 15 16 17 02 01 168 171 173 175 172 170 167 182 320 321 475
70	设防节点	3 4 5 6 10 15 14 41 40 63 620 173 231 168 243 248 253 273 370 371 488 549 561
	围捕平台	3 4 5 6 10 15 14 17 02 01 168 173 171 175 172 167 169 170 320 321 482 481 475
80	设防节点	3 4 10 15 14 41 40 62 64 233 231 168 230 248 250 254 273 349 370 488 549 558 562
	围捕平台	3 4 10 15 14 17 02 168 01 173 171 175 172 167 182 169 170 320 321 482 481 475 480
90	设防节点	3 4 10 14 41 40 85 64 171 173 190 229 168 248 250 254 273 349 369 488 549 558 562
	围捕平台	3 4 10 14 17 02 20 01 171 173 168 172 175 167 182 169 170 320 321 482 481 475 480

7　模型评价 (模型 2.2)

优点：根据犯罪嫌疑人被发现之前所能走到的所有最远点建立集合，并从这个集合出发搜索所有的“延生”节点，并通过延生出的结点与最近警力点的距离和犯罪嫌疑人距离的差距来找到封锁点，这样所得到的点既可以满足题目给出的要求，且程序设计相对容易.

缺点：可能存在警力点并非最近的调配才是最佳的情况，因为可能会出现更远的警力点到相对远的节点的可能，这样时间会很长．而本身如果将这个警力点分配到相对远的节点而更远的警力点到就近的节点，时间会更短，本文未将其考虑在内.

8　附录 2.2 程序[1,2]

```
clear; close all; clc
zuo=xlsread('C:\Documents and
Settings\Administrator\桌面\cumcm.xls',1,'B2:C583');
lv=xlsread('C:\Documents and
Settings\Administrator\桌面\cumcm.xls',2,'A2:B929');
dist=zeros(1,1);k=1;
for i=1:928
      dist(1,k)=sqrt((zuo(lv(i,1),1)-zuo(lv(i,2),1))^2
      +(zuo(lv(i,1),2)-zuo(lv(i,2),2))^2);
      k=k+1;
end
dist=dist/10;lu=zeros(582,582);k=1;
for i=1:928
      lu(lv(i,1),lv(i,2))=dist(k);
      lu(lv(i,2),lv(i,1))=dist(k);
      k=k+1;
end
for i=1:582
   for j=1:582
      if lu(i,j)==0
          lu(i,j)=10000;
      end
   end
end
for i=1:582
      lu(i,i)=0;
end
weight=lu;
for k=1:582
   for i=1:582
      for j=1:582
          if weight(i,j)>weight(i,k)+weight(k,j)
```

```
                    weight(i,j)=weight(i,k)+weight(k,j);
               end
          end
     end
end
tic  speed=60;        %小偷速度设定(km/h)
station=xlsread('C:\Documents and
Settings\Administrator\桌面\cumcm.xls',3,'B2:B81');
numstation=length(station);  %警力点个数
sp=speed/60;  numA=0;
for i=1:582      %确定无法封锁的节点集合A的初值（P点出发三分钟之内的点）
   if(weight(32,i)<(3*sp))
      numA=numA+1;
      A(numA)=i;
   end
end
fail=1;times=0;
while((numA<583)&&(fail~=0))
   times=times+1;
   fail=0;numB=0;numD=0;
   B=0;D=0;Dstation=0;
   station(:,2)=1;
  for i=1:928
   if((~isempty(find(A==lv(i,1), 1)))&&(isempty(find(B==lv(i,2), 1)))
    &&(isempty(find(A==lv(i,2), 1))))
      numB=numB+1;
      B(numB)=lv(i,2);
   end
   if((~isempty(find(A==lv(i,2), 1)))&&(isempty(find(B==lv(i,1), 1)))
    &&(isempty(find(A==lv(i,1), 1))))
      numB=numB+1;
      B(numB)=lv(i,1);
   end
end
% B=B(2:numB);  %B集合为A集合的相邻节点构成的集合
% numB=numB-1;
for i=1:numB;
   Ddistance=0;
   d1=weight(32,B(i))/sp;%小偷到该节点的最短时间
   for j=1:numstation
      d2(j,1)=weight(station(j,1),B(i));%各警力点到该节点的最短时间
   end
```

```
    D=find((d1-d2).*station(:,2)>3);
    if(isempty(D))
        numA=numA+1; A(numA)=B(i);
        B=[B(1:i-1),B(i+1:numB)];
        fail=1;  break;
    else
        numD=numD+1;
        for k=1:length(D)
            Ddistance(k)=d2(D(k),1);
        end
        [Dmin,numDmin]=min(Ddistance);
        Dstation(numD)=station(D(numDmin),1);
        station(D(numDmin),2)=0;
    end
end
end
toc  display(A)  display(B)  C=[A,B]
```

参考文献：

[1] 张志涌. Matlab教程[M]. 北京：北京航空航天大学出版社，2006
Zhang Z Y. Matlab Tutorial[M]. Beijing: Beihang University Press, 2006
[2] 王文波. 数学建模及其基础知识详解[M]. 武汉：武汉大学出版社，2006
Wang W B. Detailed Annotation of Mathematical Modeling and Basic Knowledge[M]. Wuhan: Wuhan University Press, 2006

Setting and Scheduling of Traffic and Patrol Police Service Platforms

LIU Jin, YANG Yun, WU Ke

(Faculty of Mathematics Statistics, Wuhan University, Wuhan 430072)

Abstract: Assuming that each police platform can only block one point, we create a dynamic optimization model for the optimal setting and scheduling problem of the traffic and patrol police service platforms under the consideration of the uncertainty of the criminal's fleeing direction. We first create a subset A which consists of the points with less than 3-minute away from the criminal's fleeing start point. Then we formulate a cut-set B which consists of the adjacent points of A, and search and allocate the adjacent police platforms to the points in cut-set B. If no police platforms can be allocated, we reformulate cut-set B again until an appropriate blocking condition tallies with the requirements. In this paper, we evaluate each nearly optimal blocking strategy when the criminal's speed is 60, 70, 80 or 90 km/h. Specifically, when the criminal's speed is 60 km/h, the fastest blocking time is 8.9 min.

Keywords: 0-1 programming; Floyd's algorithm; platform scheduling; best block way

第28卷 增刊1　　工 程 数 学 学 报　　Vol. 28 Supp.1
2011年11月　　CHINESE JOURNAL OF ENGINEERING MATHEMATICS　　Nov. 2011

文章编号:1005-3085(2011)07-0192-11

天然肠衣搭配优化问题的模型和计算

陆立强[1]， 俞冰清[1]， 范 岸[2]

(1- 复旦大学数学科学学院，上海 200433； 2- 中国银联股份有限公司，上海 200135)

摘 要: 本文介绍了肠衣搭配优化问题的背景、问题的分析、整理、归纳过程，讨论了两种解题思路以及相应的数学模型、计算方法以及数值结果.

关键词: 肠衣搭配；整数规划；数学建模

分类号: AMS(2000) 90B90; 90C29; 90C59　　**中图分类号:** O221.1　　**文献标识码:** A

1 问题的背景

天然肠衣(以下简称：肠衣)指的是家畜的大、小肠经刮制而成的畜产品，主要用于香肠、灌肠等食品的外衣．中国加工肠衣历史悠久，是肠衣生产大国，其产量占世界总产量的三分之一，其中约80%出口，年出口量达30多万桶，出口金额为4亿美元左右．欧洲是我国肠衣的主要出口对象，目前50%左右出口到欧洲，欧洲市场70%左右的肠衣来自于中国．我国肠衣还出口到美国、加拿大、日本、俄罗斯等国家和地区．自1985年对肠衣生产企业实施出口食品卫生注册管理以来，我国出口肠衣生产企业的整体管理水平得到了很大的提高．我国现有肠衣出口注册企业200家左右，其中对欧盟注册的就有119家左右，有力地保证了我国肠衣的对外出口．近几年，国内市场对肠衣的需求也呈递增趋势，对于肠衣生产加工企业，无论是国际市场，还是国内市场，需求均在不断地扩大，商业机会越来越多，竞争也更加激烈.

肠衣加工过程中，首先需要清洗原材料．由于化学、物理作用，难免会导致材料破损，因而必须将破损的部分剪除，这是造成原材料长度不一的主要原因；第二步是捆扎，即将清洗干净的肠衣根据一定的规格捆扎出成品．一般要求每捆成品中肠衣的总长度是固定的，所以捆扎过程中必须首先测量每一根原材料长度，然后将长短不一的原材料搭配成捆.

由于家畜肠子的湿、软、滑等特性，目前肠衣加工主要依靠人工．其中捆扎环节尤其要求工人眼明手快，所谓眼明是指原料丈量比较准确，手快则是可以快速判断手中的原料是否可以和已测量过的原料捆扎成一捆成品．符合这样要求的员工一方面年龄不能太大(否则速度太慢，影响加工速度，导致原料报废)，另一方面还得具备一定的经验(否则，无法将长短不一的原料快速捆扎成捆)．伴随着行业竞争的加剧，人工成本的增加是一种必然的趋势.

作为一种食品，一般都是原料当天进工厂，当天就要加工完成出口，不允许将剩余的原材料留作以后使用，因此对于原料的使用率有比较高的要求．与此相对的是，人工搭配时工人一般不作整体考虑，只是根据手中现有的部分原料，凭经验和直觉判断是否可以搭配成一捆成品，不可能考虑到所有的情况，无法保证原材料的充分利用.

所以无论从降低工人劳动强度与压缩人力成本，还是从提高原料产出效率的角度，肠衣加工企业亟盼依靠先进技术进行加工工艺创新，以提高劳动生产率，增强企业的核心竞争力.

2　问题的提出

某在国内具有一定影响的肠衣加工企业希望能够借助计算机计算速度快、结果可靠的特点，开发一套计算机软件，该软件只需一线工人将测量所得原料数据输入电脑，就能自动生成经过优化后的满足成品规格要求的搭配方案，这样既可以减少劳动强度、又能提高原料使用率. 企业负责人有海外留学背景，认识到这个程序的核心是要建立原料搭配优化的数学模型，因此和本文作者进行了接触.

在首次接触中，我们的第一反映是这个问题最终需要归纳成一个优化问题[1]，具体是何种问题则需要作进一步的了解. 为此我们需了解两种信息：原料信息和成品信息.

肠衣原料不同于一般的刚体，其具有一定的伸缩性，企业的测量以0.5米为一档，如：3.1米–3.5米按3米计算，3.6米–4米按3.5米计算，其余的依此类推(这样测量的结果是每一根原料的测量长度均小于等于其实际长度，企业实际上是以部分原料的免费奉送来保证成品的总长度的合格). 以下为某批次原料的测量结果，见表1.

表1:　原料描述表

长度(米)	3.1–3.5	3.6–4	4.1–4.5	4.6–5	5.1–5.5	5.6–6	6.1–6.5	6.6–7
根数	43	59	39	41	27	28	34	21
长度(米)	7.1–7.5	7.6–8	8.1–8.5	8.6–9	9.1–9.5	9.6–10	10.1–10.5	10.6–11
根数	24	24	20	25	21	23	21	18
长度(米)	11.1–11.5	11.6–12	12.1–12.5	12.6–13	13.1–13.5	13.6–14	14.1–14.5	14.6–15
根数	31	23	22	59	18	25	35	29
长度(米)	15.1–15.5	15.6–16	16.1–16.5	16.6–17	17.1–17.5	17.6–18	18.1–18.5	18.6–19
根数	30	42	28	42	45	49	50	64
长度(米)	19.1–19.5	19.6–20	20.1–20.5	20.6–21	21.1–21.5	21.6–22	22.1–22.5	22.6–23
根数	52	63	49	35	27	16	12	2
长度(米)	23.1–23.5	23.6–24	24.1–24.5	24.6–25	25–25.5	25.6–26		
根数	0	6	0	0	0	1		

对于同一批原料，成品一般分成三种规格，每种规格用4个参数加以描述(最短原料长度，最长原料长度，原料根数，总长度)，表2为对应于该批次原料的成品规格表. 其中第三种大规格的最长原材料的长度记为∞，表示最长原材料的长度没有上限.

表2:　成品规格表

类别	最小长度(米)	最大长度(米)	根数	总长度(米)
小规格	3	6.5	20	89
中规格	7	13.5	8	89
大规格	14	∞	5	89

基于上述信息，我们不难得到：

目标1 对于给定的一批原料，装出的成品捆数越多越好.

直觉告诉我们：一捆只包含1根长度为89米原料的成品和另一捆由89根长度为1米的原料组成的成品相比，前者的价值明显要高. 因此，企业提出了目标2.

目标2 对于成品捆数相同的方案，最短长度最长的成品越多，方案越好.

为了保证食品原料的新鲜，搭配方案的生成时间必须足够快，经过操作部门的沟通，得到目标3.

目标3 为了食品保鲜，要在30分钟内产生方案.

从数学角度看，提高原料使用率的本质是扩大问题可行域，也就是放宽成品规格表各档规格对于原料的限制. 其中，原料根数和总长度可以做适当的放松，即从等式约束改为不等式，于是有：

条件1 总长度允许误差范围为[−0.5, 0.5]，总根数允许误差范围为[−1, 0].

最小长度关系到成品质量，只能提高不能下降，但是提高会降低原料使用率，所以只好维持不变；作为用户，只要最小长度不减，希望最大长度越大越好.

条件2 某种规格对应原料如果出现剩余，可以降级使用.

如长度为14米的原料可以和长度大于7米的进行捆扎，成品属于7米–13.5米规格.

3 问题的分析

首先看目标2，它的处理有两个难点：第一，表述上似乎是一个最小最大问题，但实际上解决却非常困难；第二，这个要求和目标1可能是相互矛盾的，也就是满足目标1的方案可能不满足目标2，反之亦然.

事实上，企业在实际生产中也已经知道了两个目标的矛盾和求解的复杂性，所以已经“聪明”地把成品规格分成大、中、小三挡，各档成品不论其最短长度如何一律平等. 其实质是将“最短长度最长的成品最多”的要求转化为“最短长度在某个值以上的成品最多”，这样：将一个从理论上非常优美但实际无法实现的最优目标转化为看似“丑陋”但却可行的优化目标.

目标1虽然简单，但由于受目标2的限制，无法直接建立关于全部原料的优化装配模型，而应该结合条件1按照三种不同规格分步进行优化. 同时，为了提高原料使用率，可以结合条件2，扩大每种规格最大长度的上限.

综合以上思考，我们给出求解肠衣装配问题的一个总体方案：

1) 根据大规格要求，求最优解；
2) 将第1)步优化后多余的原料纳入中规格，求最优解；
3) 将第2)步优化后多余的原料纳入小规格求最优解；
4) 如果多余的原料总长小于88.5米，则优化成功.

以上总体方案的特点是：用三次优化实现整体的优化：每次优化的数学模型、求解方式相同；不同的是每一步优化的原始数据：原料长度、根数和成品规格. 由此带来的好处是：只需编写一个函数，反复调用三次，中间只需对原料、成品等数据作必要的调整.

4 模型一

4.1 搭配方式模型

对于一种给定的规格，我们用min表示原材料的最短长度，max表示原材料的最长长度，number表示最多使用的原材料根数，length表示成品长度. 由于原材料的长度是以0.5递

增的，那么，在 min 和 max 之间就总共有 $N = (\max - \min) \div 0.5 + 1$ 种不同长度的原材料，它们的长度记为 $l_1, l_2, \cdots, l_N$，对应的原料根数记为 $y_1, y_2, \cdots, y_N$.

搭配方式可以表示为 $(n_1, n_2, \cdots, n_N)$，n_i 表示该种搭配方式中长度为 l_i 的原材料根数. 因为原材料根数为 number 或 number − 1，所以

$$\text{number} - 1 \le \sum_{i=1}^{N} n_i \le \text{number}.$$

同时，由于成品的长度为 length ∓ 0.5 或 length，即

$$\text{length} - 0.5 \le \sum_{i=1}^{N} l_i n_i \le \text{length} + 0.5.$$

结合实际原料的限制，就可以得到该规格所有搭配方式 $(n_1, n_2, \cdots, n_N)$ 所要满足的条件

$$\begin{cases} \text{number} - 1 \le \sum\limits_{i=1}^{N} n_i \le \text{number}, \\ \text{length} - 0.5 \le \sum\limits_{i=1}^{N} l_i n_i \le \text{length} + 0.5, \\ 0 \le n_i \le y_i, \quad i = 1, 2, \cdots, N. \end{cases} \tag{1}$$

由于 n_i 是非负整数，并且有上界，所以满足上述不定方程组的非负整数解的个数是有限的. 利用计算机可快速求解满足上述方程组的非负整数解，每个解对应一种搭配方式.

4.2　最优搭配模型

记不定方程组解的个数为 M，其中第 j 个解为 $(a_{1j}, a_{2j}, \cdots, a_{Nj})^T$, $j = 1, 2, \cdots, M$，最优的搭配方案中第 j 种搭配方式有 x_j 捆，那么，第 j 种搭配方式消耗的第 i 种原材料为 $a_{ij}x_j$, $j = 1, 2, \cdots, N$，最优方案将总共消耗第 i 种原材料 $\sum\limits_{j=1}^{M} a_{ij}x_j$ 根，而第 i 种原材料只有 y_i 根，所以消耗的根数不可能大于原材料数，即

$$\sum_{j=1}^{M} a_{ij}x_j \le y_i \tag{2}$$

所以

$$\begin{cases} \sum\limits_{j=1}^{M} a_{1j}x_j \le y_1, \\ \qquad \cdots, \\ \sum\limits_{j=1}^{M} a_{Nj}x_j \le y_N, \end{cases} \tag{3}$$

即

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1M} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & a_{N3} & \cdots & a_{NM} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{pmatrix} \le \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}. \tag{4}$$

记

$$\overline{x}=\begin{pmatrix} x_1\\ x_2\\ \vdots\\ x_M \end{pmatrix},\quad \overline{y}=\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_N \end{pmatrix},\quad A=\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1M}\\ a_{21} & a_{22} & a_{23} & \cdots & a_{2M}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ a_{N1} & a_{N2} & a_{N3} & \cdots & a_{NM} \end{pmatrix},$$

则上式可写为

$$A\overline{x}\leq\overline{y}. \tag{5}$$

这样就确定了关于x_j的约束条件.

工厂的目标是能够得到尽量多的成品或使剩余的原材料尽量少. 先考虑得到尽量多成品的要求. 为了得到尽量多的成品，须每种小规格的捆数之和要最大，也就是$\sum\limits_{j=1}^{M}x_j$要最大. 这自然构成目标函数. 又因为实际要求，x_j为非负整数，这样就得到一个整数规划如下

$$\max\ \sum_{j=1}^{M}x_j \tag{6}$$

$$\begin{cases} A\overline{x}\leq\overline{y},\\ \overline{x}\ 为非负整数向量. \end{cases}$$

再考虑使剩余的原材料尽量少. 记N阶列向量

$$\overline{b}=\begin{pmatrix} b_1\\ b_2\\ \vdots\\ b_N \end{pmatrix},$$

使得$A\overline{x}+\overline{b}=\overline{y}$，则$b_i$就是第$i$种原材料的剩余长度. 要使剩余的原材料尽量少，即要使$\sum\limits_{i=1}^{M}b_i$尽量小. 这样得模型二

$$\min\ \sum_{i=1}^{M}b_i \tag{7}$$

$$\begin{cases} A\overline{x}+\overline{b}\leq\overline{y},\\ \overline{x},\overline{b}\ 为非负整数向量. \end{cases}$$

因为原材料的总长度是一定的，设为C. 总长度分为两部分，一部分用来做成品，消耗$89\sum\limits_{j=1}^{M}x_j$，另一部分被剩余，剩余的总长度为$\sum\limits_{i=1}^{M}b_i$，则可得

$$89\sum_{j=1}^{M}x_j+\sum_{i=1}^{M}b_j=C.$$

所以当 $\sum_{j=1}^{M} x_j$ 取到最大值时，$\sum_{i=1}^{M} b_i$ 也取到最小值．因此，(6) 与 (7) 是等价的．

由于 (6) 共有 M 个决策变量，而 (7) 的决策变量是 $M-N$ 个，从理论上来讲，(6) 比 (7) 的求解速度要快，因此，通常采用前者来进行求解．但实际上，由于 M 比 N 大很多，因此，M 和 $M-N$ 相差也不是很大，两者的计算速度也差异不大．

4.3 搭配方式模型求解

由于厂家的成品规格对应原料档次数不是常数，搭配方式的生成无法使用循环方式来完成，因而必须通过递归处理．而且原材料的长度是以0.5递增的，为了加快运算速度，可以把长度都乘以2转换成整数来处理．那么，相邻两种长度的肠衣的长度之差为1．

如果采用Matlab编程，递归所需运行时间较长，很可能导致总体计算时间超过30分钟的限制，因此一般采用C程序生成小规格，这样生成搭配方式可以在几分钟甚至十几秒内完成．

4.4 优化模型求解

本文采用Lingo来计算整数规划[2,3]．搭配方式保存在txt文件中．虽然Lingo可以导入txt文件，但其导入时是分行导入，而对于M比较大的情况时，分行导入显然是不合适的．若直接在Lingo上输入小规格情况数据也是不现实的．因为Lingo可以直接导入Excel文件，因此，本文先将txt文件在Excel中打开，并加入原材料根数的数据，再定义两个名称，保存为Excel文件之后，导入到Lingo中．同时，在算完整数规划之后，本文还将最优解输入到txt文件中，以方便查阅．

第一步　下面是求解大规格搭配方案整数规划的Lingo代码．

```
model:
sets:
rows/1..2862/:x;
cols/1..23/:y;
table(rows,cols):A;
endsets
max=@sum(rows(i):x(i));
@for(cols(j):
@sum(rows(i):A(i,j)*x(i)) ¡= y(j));
@for(rows(i):@gin(x(i)));
data:
A=@ole('mydata3.xlsx','kinds');
y=@ole('mydata3.xlsx','y');
@text('mysol3.txt')=rows '最优解' x;
enddata
end
```

用上面给出的Lingo代码，在 $M=2862$ 的情况下，最优目标值为137，最优解如表3所示．表中每行表示一种搭配方式，行中最后一个数字表示该种搭配对应的捆数．

查看剩余，只剩余一根长度为18的，因此这样的组装方案显然是最优的．

第二步　将上面多余的原料降级到中规格进行搭配，可算得 $M=19635$，最优捆数为37捆，详细情况如表4所示．剩余原料如表5所示．

尽管剩余材料总长度超过89.5，但是用5.1所述捆数上限模型可以算得中规格捆数的上限为37捆，和本算例所得相等．

表3: 大规格搭配方案

	原料长度(米)																								捆数
	14	14.5	15	15.5	16	16.5	17	17.5	18	18.5	19	19.5	20	20.5	21	21.5	22	22.5	23	23.5	24	24.5	25	25.5	
	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	4
	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
	1	0	0	0	0	1	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
	1	0	0	0	0	1	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	7
	1	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	8
搭	0	1	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	15
	0	1	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
配	0	0	1	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	8
	0	0	1	0	1	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	16
方	0	0	1	0	0	1	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	6
	0	0	0	1	1	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	10
式	0	0	0	1	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	10
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	1	0	0	0	0	6
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	0	0	0	0	0	0	2
	1	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	13
	0	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	17
	0	0	0	1	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	1	1

表4: 中规格搭配方案

	原料长度(米)															捆数
	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5	18	
	0	0	1	1	0	0	0	0	3	0	0	0	3	0	0	2
	0	0	1	1	0	0	0	0	0	3	0	3	0	0	0	2
捆	0	0	1	0	3	0	0	0	0	0	0	0	1	3	0	3
	0	0	1	0	0	0	4	0	0	0	0	0	0	3	0	1
扎	0	0	1	0	0	0	0	1	2	0	4	0	0	0	0	4
	0	0	0	2	1	0	0	0	0	0	2	2	0	1	0	2
方	0	0	0	1	1	0	0	2	0	2	0	0	0	2	0	5
	0	0	0	1	0	2	0	0	1	0	0	4	0	0	0	11
式	0	0	0	0	4	0	0	0	0	3	0	0	0	0	1	1
	0	0	0	0	1	0	0	4	0	0	0	3	0	0	0	1
	0	0	0	0	0	0	4	0	1	1	0	0	2	0	0	4
	0	0	0	1	0	0	1	0	2	0	2	2	0	0	0	1

第三步 将上述多余原料纳入小规格原料进行优化搭配时，即使不考虑长度和根数的变化区间，只考虑严格相等约束条件，M 也近500万个解，难以想象在微机上可以实施计算，必须将求解理论最优转化为求可行的最优(次最优解)，对于企业也是可以接受的. 得到次最优解的

主要方式是缩小搜索空间，如我们可以等步长地从 M 个搭配方式中选取Lingo所允许的最大个数作为搜索子空间进行优化．对于本例数据，我们得到需用Lingo作优化的24564个搭配方式，最终得到目标函数值为16，如表6所示．剩余原料情况如表7所示．

尽管剩余材料总长度超过89.5，但是用5.1所述捆数上限模型可以算得小规格捆数的上限为16捆，所以所得近似最优可以认为是全局最优的．

表5:　中规格原料剩余情况

长度(米)	7	7.5	8	9.5	13	13.5
根数	24	24	8	1	1	1

表6:　小规格近似优化搭配方案

	原料长度(米)														捆数
	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	9.5	13	13.5	
	4	0	1	1	4	4	1	0	1	3	0	0	1	0	1
	3	4	4	1	0	2	0	4	0	0	1	1	0	0	1
搭	2	4	3	0	4	0	0	4	1	1	1	0	0	0	1
	2	3	2	4	2	0	0	4	3	0	0	0	0	0	1
配	3	1	0	2	3	4	0	3	1	1	0	0	0	2	
	1	3	7	0	1	3	2	0	1	0	1	0	0	1	1
方	7	2	3	3	0	3	1	0	1	0	0	0	0	3	
	0	4	3	8	1	1	1	1	0	0	1	0	0	0	3
式	4	0	0	0	8	0	1	6	1	0	0	0	0	1	
	0	1	4	1	0	0	4	2	0	8	0	0	0	0	1
	0	0	1	1	0	0	7	0	5	6	0	0	0	0	1

表7:　小规格原料剩余情况

长度(米)	3	3.5	5.5	7
根数	27	1	2	1

5　模型二

模型一先求出符合条件的所有搭配方式，再根据这些方式展开全局优化，所得到的最优解一定也是实际需要的最优解，其不足在于：在进行小规格优化搭配时，因为搭配方式过多而只能得到局部极优解．

事实上，因为同一批次原料根数不会很大，因此成品数量远远小于可供选择的搭配方式数量．由此我们得到解决问题的另外一种思路：模仿人工搭配方式，将最优捆数和搭配方式一起作为优化变量进行求解．具体做法也分两种：

1) 先估计出成品的上限，再求出可能的搭配方式；

2) 直接求出最优捆数和搭配方式.

5.1 捆数上限模型

对符合某种成品规格的原料，设number, length, N, l_j, y_j, $j=1,2,\cdots,N$ 和模型一的意义相同；记 M 为该规格成品数量，x_j 为这些成品中第 j 种原料被使用的总数 $(j=1,2,\cdots,N)$，则 M 满足

$$
\max\ M \tag{8}
$$

$$
\begin{cases}
M(\text{length}-0.5)\le \sum\limits_{j=1}^{N} x_j l_j \le M(\text{length}+0.5),\\
M(\text{number}-1)\le \sum\limits_{j=1}^{N} x_j \le M\cdot \text{number},\\
x_j\le y_j,\quad j=1,2,\cdots,N,\\
M,\ x_j\ \text{为非负整数}.
\end{cases}
$$

模型(8)参数意义明确，个数也较少，可以直接调用Lingo软件求解得到最优解 M_0，进一步，设 x_{ij} 为第 i 捆成品中第 j 种原料的根数 $(i=1,2,\cdots,M_0,\ j=1,2,\cdots,N)$，则成立

$$
\begin{cases}
\text{length}-0.5\le \sum\limits_{j=1}^{N} x_{ij} l_j \le \text{length}+0.5,\quad i=1,2,\cdots,M,\\
\text{number}-1\le \sum\limits_{j=1}^{N} x_{ij} \le \text{number},\\
\sum\limits_{i=1}^{M_0} x_{ij}\le y_j,\quad j=1,2,\cdots,N,\\
x_{ij}\ \text{为非负整数}.
\end{cases} \tag{9}
$$

5.2 捆数和捆扎方案联合求解模型

设：number, length, N, l_j, $y_j\ (j=1,2,\cdots,N)$ 和模型一的意义相同；记：M 为成品数量，x_{ij} 为第 i 捆成品中第 j 种原料的根数 $(i=1,2,\cdots,M,\ j=1,2,\cdots,N)$，则 M 和 x_{ij} 满足

$$
\max\ M \tag{10}
$$

$$
\begin{cases}
\text{length}-0.5\le \sum\limits_{j=1}^{N} x_{ij} l_j \le \text{length}+0.5,\quad i=1,2,\cdots,M_0,\\
\text{number}-1\le \sum\limits_{j=1}^{N} x_{ij} \le \text{number},\\
\sum\limits_{i=1}^{M} x_{ij}\le y_j,\quad j=1,2,\cdots,N,\\
M,\ x_{ij}\ \text{为非负整数},
\end{cases}
$$

6 模型分析

模型(8)的变量个数为 $M_0\times N$，比模型(6)的变量个数 M 相比要小得多，因此求解相对比较容易，一般不会发生因为 M 太大而无法优化的情况．但是，在求解搭配方案问题(9)时，因为没有目标函数，所以往往每次只能得到局部极优.

模型(10)可以直接通过Lingo求解，但因为第三个约束条件的求和项数也是一个优化变量，因此本质上它是一个非标准的整数规划问题，即使用Lingo也无法得到整体最优，每次只能得到一个局部极优.

更值得指出的是，在依据(9),(10)求解搭配方案的过程中，一次优化只能得到一种搭配方式的最多捆数，为了得到全部材料的搭配，需要修改剩余原料数据再次优化，如此逐步进行，直到剩余材料无法成捆为止.

总体而言，模型二和人工搭配方式比较接近，结果也容易理解，但执行过程人工干预较多，一般不能保证在30分钟内得到搭配方案．模型一对问题的理解比较彻底，是一个标准的整数规划模型，理论上可以得到真正的最优解，只要编程得当，基本不需要人工干预，符合企业的最终要求．不过由于该问题是NP困难的，所以在搭配方式较多的情况下，理论最优解无法在规定时间内求得，必须通过子空间内搜索而得到一个局部极优解.

7　总结

作者之所以将此问题公开作为全国数模竞赛题，是因为无论是通过模型一还是模型二，这个问题可以表达为一个数学规划问题，而这对于一个接受过数学建模训练的大学生而言应该不是一个难题．但是，实际情况没有想象的那样乐观.

首先，相对C题，选做本题的同学比较少．其中可能有两个原因，第一，学生已经习惯于通过网络搜集解题信息(这是当代大学生应该具备的技能)，但是“肠衣搭配优化问题”是一个源于中国的“土”问题，网络上除了肠衣厂商的广告网站，几乎得不到任何有用的资料，许多参赛同学因此心存怯意，不敢尝试．第二，也有一些同学感觉它和经典的“钢材切割问题”非常相像，只不过后者是化整为零，前者是集零成整，应该容易求解．但是，这表面的细微差别却导致问题的本质发生了变化，使变量个数成倍增加，以致无法在规定时间内用现有的软件完成计算，他(她)们也因此而放弃了努力．对于第二类同学，作者认为在数学建模竞赛过程中，模型固然是一个核心，但是有了模型如何求解，也是解决实际问题的一个重要环节．在计算机软硬件条件十分成熟的今天，我们当然不必事必躬亲，自己设计算法，自己编程求解，而应采用充分利用国内外先进的计算软件．但是，实际问题中会出现计算规模超过计算机能力的情况，一种可行的解决办法就是将漂亮的理论最优解代之以实用的局部极优解．从此次阅卷的情况来看，D题获奖的参赛队，都或多或少地应用了这个思想．这也提醒了我们教师在教学过程中应该适当增加近似计算方面的内容.

其次，本次做D题的绝大多数参赛队采用的建模思路是第二种模型，没有用模型一．其中原因我们在前面也曾分析过：它的思路和人工操作比较接近，有的参赛队甚至在论文中指出曾向生产一线的工人请教过．相比之下，模型一思路不是那么直接，虽然它是经典的整数规划模型，所得解的可信度更高，也更加容易改编成企业需要的软件产品．因此，通过本题，参赛同学会对整数规划模型及其计算有更深层次的理解，这对于将来解决实际问题是有帮助的.

最后，要特别感谢谭永基、叶其孝、蔡志杰等教授在2010年重庆工业应用数学研讨会上的启发和帮助，也要感谢全国大学生数学建模竞赛组委会及专家组在出题过程中所给予的指导.

参考文献：

[1] 胡运权. 运筹学(第3版)[M]. 北京：清华大学出版社，2005
Hu Y Q. Operations Research (3rd Edition)[M]. Beijing: Tsinghua University Press, 2005

[2] 刘璟忠. 基于Lingo语言求解组合优化问题[J]. 现代计算机(专业版), 2005, (6): 81-82
Liu J Z. Solving optimization of combination problem based on Lingo[J]. Modern Computer, 2005, (6): 81-82
[3] 丁小东, 姚志刚, 程高. Lingo语言与0-1混合整数规划衔模型的再结合[J]. 物流工程与管理, 2009, 31(10): 72-75
Ding X D, Yao Z G, Cheng G. Lingo language in the logistics distribution center location[J]. Logistics Engineering and Management, 2009, 31(10): 72-75

Natural Casing Assembling Optimization: Models and Computation

LU Li-qiang[1], YU Bing-qing[1], FAN An[2]

(1- School of Mathematical Sciences, Fudan University, Shanghai 200433;
2- China Unionpay Co., Ltd, Shanghai 200135)

Abstract: This paper discusses the problem of optimization in natural casing assembly. Based on the thorough study on its background and requirements, we propose two mathematical models and analyze their computing behavior in integer programming.

Keywords: natural casings assembling; integer programming; mathematical modeling

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0203-08

一类天然肠衣原料搭配优化问题的启发式求解

杨 忠, 张岐良, 徐小辉

指导教师: 凌巍炜, 李繁春

(江西应用技术职业学院基础部, 赣州 341000)

编者按: 本文基于对问题基本要求(1)的理解，给出了对整体(所有原料段)进行优化搭配的模型，作者同时指出了这个模型解法的NP问题，为此，应用了“改进的领域整点搜索的启发式算法”，在规定时间内给出了求解方案．当然，本文存在明显的不足是启发式算法得到的捆数超过了理论上限，这可能是降级处理不当所致．

摘 要: 本文针对肠衣搭配问题建立了优化模型，给出了相应启发式算法，较完善的解决了本问题．针对要求(1)，本文将问题分解为分别求规格1、规格2和规格3装出成品捆数最多的问题．在3种规格成品的根数和总长度都固定不变且每种原料长度和根数都已知的条件下，基于规划理论，对各规格用于组装的原料根数和长度进行了约束，进而建立了优化模型，并求出组装出的成品最大捆数之和为182捆．针对要求(2)，基于各规格成品组装捆数最多的前提下，对相同最大捆数下的方案，通过Lingo软件求得了每个方案中每捆最短的原料和原料数量，通过比较求得每种规格取最短原料中最长的数量最多的方案．针对要求(3)、(4)，首先，对在不降级组装条件下，重新对用于组装的原料根数和长度进行约束，继而求出了不降级组装时各规格组装出的捆数；然后，在假设剩余原料只能降一个等级的条件下，得到放宽条件下原料降级组装的捆数；最后，由放宽条件下的组装成品捆数和降级组装的成品捆数求得了组装的成品捆数最多为208捆及每种规格取最短原料中最长的数量最多的方案．针对时间限制的要求(5)，在建立的整数规划模型求解规模较大，编程运行时间超过30分钟的情况下，本文创造性的引入了一个启发式算法，并由此算法求得了上述4个要求的近似解．由启发式算法编写的程序可在1118秒内得到结果，即求得满足所有5个要求的较优搭配方案，工人可以根据这个方案“照方抓药”进行生产．同时本文还给出了改进的邻域整点搜索算法求解本问题，对两种启发式算法的结果和复杂度进行了分析比较．

关键词: 肠衣搭配；优化模型；启发式算法

分类号: AMS(2000) 90C05 中图分类号: O221.4 文献标识码: A

1 问题分析

在给出了对应3种成品规格和原料数据情况下，公司对要设计的搭配方案提出了5个要求．基于这5个要求可知，问题是要在组装出的成品捆数最多的前提下，再对相同最大捆数的搭配方案进行比较，最终选出最好的搭配方案．对于要求(1),(2)，因为每种规格成品在捆数最多的条件下，都可能存在多种方案，所以在这里应该分别求出规格成品(1)、规格成品(2)和规格成品(3)的最大捆数．由于每种规格成品可能在最大捆数下有多种搭配方案，在这种情况下可以分别列出3种规格的每个方案中每捆长度最短的原料，再算出每种方案中最短长度最大的原料数量，那么此数量最多的方案也就是最好的方案．对于要求(3),(4)，是为了提高原料使用率，也就是对3种成品规格的根数、总长度和每种成品规格的原料长度区间增加了范围．因为对每种产品规格的根数、总长度和原料长度区间增加范围，显然装出的成品捆数会比没有增加范围前的捆数多．在每种规格成品捆数变多情况下，可以再对每种规格的成品进行要求(1),(2)的处理，那么最终求出的方案即为最好的．对于要求(5)，为了食品保鲜，要求在30分钟内产生方案，在这里可以考虑是否对肠衣如何搭配给出一种启发式算法，并使这种算法编写的程序运行时间不超过30分钟，那么对于此问题就归结于有效算法的构造了．

2 模型假设

1) 假设原料最多只能降一级使用；

2) 假设每档长度的原料都以0.5米为一档，如：3–3.5米，按3米计算；

3) 假设每次肠衣制作加工，分割成的原料对应长度与根数统计是正确的；

4) 假设本次的肠衣制作加工完剩余的原料不能用于下次的组装.

3 符号说明

m_i: 表示第i档长度原料的长度，$i=1,2,\cdots,46$；

x_{ij}: 表示第j捆成品中第i档长度原料的根数，$i=1,2,\cdots,46$, $j=1,2,\cdots,n$；

p_i: 表示第i档长度原料的根数；

L: 表示每捆成品的总长度；

s_u: 捆法的方案；

y_k: 表示3米到第k规格最长长度原料的长度档数；

v_i: 表示原料不降级使用时第i档长度原料组装完后剩余的根数；

g_k: 表示第k规格成品中每捆的根数，$k=1,2,3$；

n_k: 表示原料不降级使用时第k规格原料组装为成品的捆数；

Q_1: 表示规格2剩余的原料降级后与规格1剩余的原料组装为成品捆数；

Q_2: 表示规格3剩余的原料降级后与规格2剩余的原料组装为成品捆数；

s_{ij}: 表示原料降级使用时第j捆成品中第i档长度原料的根数；

M: 表示最短长度最长的成品捆数.

4 模型建立与求解

4.1 模型的建立

考虑分步建立模型的办法，首先建立以3种规格产品组装的最多捆数为目标，即建立满足要求(1),(3),(4)的优化模型；其次，基于第一步模型求解结果，建立满足要求(2)的优化模型.

4.1.1 基于要求(1),(3),(4)的模型建立

为了提高原料的使用率，每捆产品的总长度允许有0.5米的误差、总根数允许比标准少1根并且原料可以降级使用(假定只能降一级，规格1对应的原料不能降级使用). 从而，以各种规格产品的最大捆数之和为目标函数，建立整数规划模型.

目标函数为

$$\max \quad Z=\sum_{k=1}^{3}n_k+Q_1+Q_2,$$

约束条件为

$$L-0.5 \le \sum_{i=y_{k-1}+1}^{y_k} m_i x_{ij} \le L+0.5, \quad k=1,2,3, \quad j=1,2,\cdots,n_k, \tag{1}$$

$$g_k-1 \le \sum_{i=y_{k-1}+1}^{y_k} x_{ij} \le g_k, \quad k=1,2,3, \quad j=1,2,\cdots,n_k, \tag{2}$$

$$p_i-\sum_{j=1}^{n_k} x_{ij}=v_i, \quad k=1,2,3, \quad j=1,2,\cdots,n_k, \tag{3}$$

$$L-0.5 \le \sum_{i=y_{k-1}+1}^{y_{k+1}} m_i s_{ij} \le L+0.5, \quad k=1,2,3, \quad j=1,2,\cdots,n_k, \tag{4}$$

$$g_k-1 \le \sum_{i=y_{k-1}+1}^{y_{k+1}} s_{ij} \le g_k, \quad k=1,2,3, \quad j=1,2,\cdots,n_k, \tag{5}$$

$$\sum_{j=1}^{Q_1} s_{ij} \le v_i, \quad i=1,2,\cdots,y_1, \tag{6}$$

$$\sum_{j=1}^{Q_1+Q_2} s_{ij} \le v_i, \quad i=y_1+1,\cdots,y_2, \tag{7}$$

$$\sum_{j=1}^{Q_2} s_{ij} \le v_i, \quad i=y_2+1,\cdots,y_3, \quad x_{ij}, s_{ij}, v_i \in Z^*, \tag{8}$$

其中约束 (1), (2) 分别表示要求 (3) 的成品总长度约束和总根数约束；约束 (3) 为每种规格的原料总根数减去对应原料组装成成品的根数等于剩余的根数；约束 (4), (5) 分别表示降级使用的原料部分满足要求 (3) 的成品总长度约束和总根数约束；约束 (6), (7), (8) 分别为每种规格产品中用于降级使用的原料数量不能超过每种规格产品中降级前组装成成品中剩余的原料数量.

4.1.2　增加考虑要求 (2) 的“优中选优”的模型建立

对于成品捆数相同的方案，最短长度最长的成品越多，方案越好，即

$$\max \operatorname{card}\{\min\{s_u\},\ u=1,2,\cdots,n,\ 且\ \min\{s_u\}=\mathrm{M}\},$$

其中 $\max\{\min\{s_u\}\}=\mathrm{M}$；$\operatorname{card}\{A\}$ 表示集合 A 的元素个数；s_u 表示捆法的方案；M 表示最短长度最长的成品捆数.

4.2　基于满足要求 (5) 的启发式算法求解

根据该模型的 NP 属性，直接求解，求解时间远远超过了题目要求 (5) 的规定时间，所以需要构造启发式算法进行求解.

4.2.1　启发式算法的构建

算法符号说明：q_j 表示第 j 步最多捆数的取法.

步骤1 初始化 $j=1$, $k=1$，求解下列模型：

$$\max \quad q_j$$
$$\begin{cases} g_k - 1 \leqslant \sum\limits_{j=y_{k-1}+1}^{y_k} q_j \leqslant g_k, \\ L - 0.5 \leqslant \sum\limits_{j=y_{k-1}+1}^{y_k} m_i q_j \leqslant L + 0.5, \\ v_i = p_i - q_j x_{ij} \geqslant 0, \\ q_j, p_i \in Z^*; \end{cases}$$

步骤2 将结果进行选择，即对最短长度最长的成品越多方案越好：

$$\max\{\min\{s_u\}\} = M, \quad \max\ \mathrm{card}\{\min\{s_u\} \mid u = 1,2,\cdots,n,\ 且\ \min\{s_u\} = M\},$$

得到一种最好的方案；

步骤3 将 v_i 换为 p_i 返回步骤1，令 $j=j+1$，当 m_j 无可行解或 $m_j=0$ 时，终止 j；执行 $n_k = \sum\limits_{i=1}^{j} m_j$，令 $k=k+1$；

步骤4 将步骤3中剩余的重新记为 p_i 再将 i 的求和范围变为 $y_{k-1}+1$ 到 y_{k+1}，转步骤1.

4.2.2 启发式算法的求解结果

根据启发式算法，求得了每种规格下的最大捆数情况下的最好方案．规格1的方案见表1；规格2的方案见表2；规格3的方案见表3；规格2降级使用的方案见表4；规格3降级使用的方案见表5.

通过启发式算法，综合考虑要求(1),(2),(3),(4)，求得规格1的捆数为13捆，规格2的为34捆，规格3的为134捆，规格2降级使用最多总捆数为6捆，规格3降级使用最多总捆数为17捆．求得总捆数为208捆．我们运行该启发式算法用时为1318秒(小于30分钟)，所以该算法满足要求(5)．并且我们求得的启发式算法的捆法方案比较少，即可操作性强．

表1: 规格1的方案

捆法	捆数	3	3.5	4	4.5	5	5.5	6	6.5
1	11	0	5	3	3	2	2	3	1
2	2	10	0	0	0	2	3	0	5
剩余		23	4	6	8	1	0	1	0

表2: 规格2的方案

捆法	捆数	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5
1	22	0	1	0	1	0	0	0	1	1	1	1	2	0	1
2	9	0	0	0	0	2	0	0	2	1	0	0	1	2	0
3	3	0	0	0	0	0	0	5	0	0	0	0	2	0	1
剩余		24	2	20	3	3	23	6	0	0	1	0	0	0	0

表3:　规格3的方案

捆法	捆数	14	14.5	15	15.5	16	16.5	17	17.5	18	18.5	19	19.5	20	20.5	21	21.5	22	22.5	23	23.5	24	24.5	25	25.5
1	49	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	32	0	0	0	0	0	1	1	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	16	0	0	0	2	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
4	12	0	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
5	6	0	0	0	1	0	1	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0
6	5	0	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
7	2	0	0	0	1	0	2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
8	1	0	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0
9	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
剩余		35	26	3	2	0	0	1	0	1	0	1	2	0	0	0	0	0	0	1	0	0	0	0	0

表4:　规格2降级使用

捆法	捆数	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	13.5
1	5	4	0	1	1	0	0	0	0	4	0	4	0	0	4	1	0	0	0	0	0	0	0
2	1	3	3	1	0	1	0	1	0	4	0	0	1	0	3	1	0	0	1	0	0	0	0
剩余		0	1	0	3	0	0	0	0	0	2	0	2	3	0	0	0	0	0	0	0	0	0

表5:　规格3降级使用

捆法	捆数	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13	⋯
1	13	0	2	2	0	0	0	0	0	0	0	0	1	0	⋯
2	3	3	0	0	0	0	0	0	0	0	0	0	0	0	⋯
3	1	0	0	0	0	0	1	0	1	0	0	1	0	0	⋯
剩余		1	1	0	0	0	0	0	0	0	0	0	0	0	
捆法	捆数	19	19.5	20	20.5	21	21.5	22	22.5	23	23.5	24	⋯	25.5	
1	13	0	1	0	0	1	0	1	0	0	0	0	⋯	0	
2	3	0	0	0	2	0	2	0	0	0	0	0	⋯	0	
3	1	0	0	0	0	0	0	0	1	0	1	0	⋯	0	
剩余		0	0	0	0	0	0	0	0	0	0	0	⋯	0	

4.3 改进的邻域整点搜索的启发式算法[1]

4.3.1 有关概念

松弛最优解：指略去变量为整数的条件原IP问题(记为B)松弛问题(记为A)的最优解，即

$$\begin{cases} x^* = (x_1^*, x_2^*, \cdots, x_n^*)^T, \\ Z^* = \sum\limits_{j=1}^{n} c_j x_j^*. \end{cases}$$

松弛最优解邻域(简称邻域)：指满足如下条件的解点 X 的集合 $[x_j^*] \le x_j \le [x_j^*] + 1$.

限界约束：分为上界约束 $x_j \le [x_j^*] + 1$ 和下界约束 $[x_j^*] \le x_j$.

弱化约束：指满足条件 $x_j' \le x_j^*$ 的约束条件 $x_j \le x_j'$ (对求极大解)，或满足 $x_j' \ge x_j^*$ 的约束条件 $x_j \ge x_j'$ (对求极小解).

4.3.2 改进的邻域整点搜索的启发式算法

$$\begin{cases} L - 0.5 \le \sum\limits_{i=y_{k-1}+1}^{y_k} m_i x_{ij} \le L + 0.5, \\ g_k - 1 \le \sum\limits_{i=y_{k-1}+1}^{y_k} x_{ij} \le g_k, \\ p_i - \sum\limits_{j=1}^{n_k} x_{ij} = v_i, \\ L - 0.5 \le \sum\limits_{i=y_{k-1}+1}^{y_{k+1}} m_i s_{ij} \le L + 0.5, \\ g_k - 1 \le \sum\limits_{i=y_{k-1}+1}^{y_{k+1}} s_{ij} \le g_k, \\ \sum\limits_{j=1}^{Q_1} s_{ij} \le v_i, \quad i = 1, 2, \cdots, y_1, \\ \sum\limits_{j=1}^{Q_1+Q_2} s_{ij} \le v_i, \quad i = y_1 + 1, \cdots, y_2, \\ \sum\limits_{j=1}^{Q_2} s_{ij} \le v_i, \quad i = y_2 + 1, \cdots, y_3, \end{cases} \tag{9}$$

$$x_{ij}, s_{ij}, v_i \in Z^*, \tag{10}$$

$$\max \operatorname{card}\{\min\{s_u\}, u = 1, 2, \cdots, n, \text{ 且 } \min\{s_u\} = \mathrm{M}\}. \tag{11}$$

步骤1 略去整数约束(10)求其松弛最优解，即先给定 n 值，n 从捆数上限取起，如第一等级从上限为14捆取起，如果无可行解，n 减少1再代入(7)式找可行解，直到找到可行解；

步骤2 求得松弛解记为 x_{ij}^*, s_{ij}^*；

步骤3 确定邻域的上下界：$[x_{ij}^*]$, $[x_{ij}^*] + 1$, $[s_{ij}^*]$, $[s_{ij}^*] + 1$；

步骤4 列出邻域边界的整点集合：

$$x_1^* = (x_{ij}^*, s_{ij}^*), \quad x_2^* = (x_{ij}^*, [s_{ij}^*] + 1), \quad x_3^* = ([x_{ij}^*] + 1, s_{ij}^*), \quad x_4^* = ([x_{ij}^*] + 1, [s_{ij}^*] + 1);$$

步骤5 可行性检验，将整点代入(6)式，找出满足的集合，通过检验就是原问题的近似最优解；

步骤6 代入(8)式取其中最好的解.

4.3.3 基于改进的邻域整点搜索的启发式算法求解结果 (略)

4.4 两种启发式算法的求解结果比较

第一种启发式算法的方案捆数见表6．改进的邻域整点搜索启发式算法的方案捆数见表7.

用第一种启发式算法求得的最多总捆数为208捆，用改进的邻域整点搜索的启发式算法求得的总捆数为166捆．从结果可看出第一种启发式算法求得的结果更接近于最优解，其算法程序运行的时间为1117秒，改进的邻域整点搜索的启发式算法程序运行的时间为600秒．对两种算法的综合评价：邻域整点搜索算法的优点为操作简便，能较快的求解结果和方案，缺点是结果只是局部极优；第一种启发式算法的优点是解的值接近于捆数的上限，缺点是操作较繁琐.

表6:　对于5种规格的总捆数

规格1的总捆数	规格2的总捆数	规格3的总捆数	规格2降级使用总捆数	规格3降级使用总捆数	3种规格总捆数
13	34	134	6	17	208

表7:　改进的邻域整点搜索的启发式算法的方案捆数

规格1的总捆数	规格2的总捆数	规格3的总捆数	规格2降级使用总捆数	规格3降级使用总捆数	3种规格总捆数
9	25	103	24	5	166

5　结果分析

此类问题是属于NP难题，不存在多项式精确的算法，只能寻求一种近似算法．通过表2可知规格1的根数为292根，最多能捆14根，余12根，总长度为1.355×10^3米，最多能组成14捆，余59.5米，则规格1上限为14捆．同理，规格2的上限为34捆，规格3上限为135捆．通过第一种启发式算法求得规格1的捆数为13捆，规格2为34捆，规格3为134捆，通过比较可知求得的结果与上限非常接近，所以该启发式算法是有效的，而且我们算法的运行用时为1118秒 (小于30分钟)，算法满足题意．并且我们求得的启发式算法的捆法方案比较少，可操作性强.

参考文献：

[1] 郭永发，等. 邻域整点搜索法求解整数规划[J]. 数学的实践与认识，2006, 36(11): 100-104
Guo Y F, et al. Neighborhood integer point search method to solve integer programming[J]. Mathematics in Practice and Theory, 2006, 36(11): 100-104

A Heuristic Method for Solving the Assignment Problem of Natural Casing Materials

YANG Zhong, ZHANG Qi-liang, XU Xiao-hui
Advisor: LING Wei-wei, LI Fan-chun

(Department of Basic Courses, Jiangxi College of Applied Technology, Ganzhou 341000)

Abstract: In this paper, we establish an optimization model for the assignment problem of the natural casing, and design a heuristic algorithm to solve the problem. For the requirement (1), we reduce the problem into a maximal number of bundles for size 1, size 2 and size 3. Based on the Planning theory, we establish an optimization model for the problem on condition that the number and the length of each product are fixed and known, and also we get the maximal number of bundle is 182. For the requirement (2), we obtain the material number in each bundle of the scheme, which gets the maximal number of bundles by employing Lingo 10.0. As compared with all schemes, we obtain the best scheme which gets the maximal number of maximal length of shortest materials in each bundle. For the requirements (3) and (4), first of all, we obtain the number of each bundle size on the assumption of no degraded using, and then we get the bundle number of remainder materials on the assumption of degraded using. Finally, we get the maximal bundle number 208 and obtain the optimal scheme. For the time limit requirement (5), based on the establishment of large-scale integer programming model, this paper introduces a creative heuristic algorithm and obtains the approximate solution. The programs of the proposed heuristic algorithm can attain results within 1118 seconds. Worker can arrange the production according to side fill a prescription. Also, we give an improved neighborhood point search algorithm to solve this question. We further compare the results and complexities of the two proposed methods.

Keywords: casing assignment; optimization models; heuristic algorithm

第28卷 增刊1 工 程 数 学 学 报 Vol. 28 Supp.1
2011年11月 CHINESE JOURNAL OF ENGINEERING MATHEMATICS Nov. 2011

文章编号:1005-3085(2011)07-0211-08

天然肠衣最大捆数的搭配问题研究

周 亚， 刘振堂， 陆雪阳

指导教师： 黄 德

(广西城市职业学院，广西 532100)

编者按：本文作者对问题的理解比较透彻，首先考虑给出符合成品规格和原料供给的搭配方案，在此基础上求得符合问题要求的最佳搭配方案．考虑到可行的搭配方案数量过多会无法在规定时间内求解结果，文章提出“同一捆中原料种类不超过三种”的想法，因而使寻优过程变得可行．当然，本文的不足也因此而来：剩余的原料集中在某些规格，会降低原料的使用率．

摘 要：本文从公司对搭配方案的具体要求出发求出构成一捆产品的全部方案．由于所得方案的总量过大，因此我们舍弃那些原料种类超过三种以上的方案，在原料不可降级使用时建立捆数最大模型．为了提高原料的使用率所以本文又在考虑原料可降级使用条件下，得出一个全部产品一起考虑的综合整数规划模型．在使得总捆数最多的优化前提下从所有可能方案中找出最优的捆绑组合方案，通过模型解得最多可以捆绑肠衣191捆，并确定了最优的搭配方案．

关键词：整数规划；分类方案；Matlab；Lingo

分类号：AMS(2000) 90C05 中图分类号：O221.1 文献标识码：A

1 问题重述 (略)

2 符号说明

第I类：	规格为3–6.5的成品；
第II类：	规格为7–13.5的成品；
第III类：	规格为14–25.5的成品；
$A_i=(a_{i1},a_{i2},\cdots,a_{i8})$：	构成一捆第I类成品的第i种方案；
$B_i=(b_{i1},b_{i2},\cdots,b_{i14})$：	构成一捆第II类成品的第i种方案；
$C_i=(c_{i1},c_{i2},\cdots,a_{i24})$：	构成一捆第III类成品的第i种方案；
$o=(o_1,o_2,\cdots,o_8)$：	第I类成品的原料数；
$p=(p_1,p_2,\cdots,p_{14})$：	第II类成品的原料数；
$q=(q_1,q_2,\cdots,q_{24})$：	第III类成品的原料数．

3 模型的假设 (略)

4 模型的建立与求解

问题给出的五个条件要想一次达到是非常困难的，故我们采用分步的方式来满足这些要求．要想知道所给原料能够构成多少捆数，首先应该找出那些能构成一捆的所有方案，然后再求最大捆数，下面我们就分这两个步骤来进行研究．

4.1 构成一捆成品的方案

要想求出符合给定条件的最大捆数，首先必须要分析构成一捆成品的方案有多少种．比如须分析对于第I类产品可以从3米、3.5米、4米、…、到6.5米中任取几个构成总数为19或20根，总长度为88.5米、89米或89.5米的方案个数．据此，为了得出构成一捆第I类成品的方案，我们建立分类模型(一)如下：

分类模型(一)

$$\begin{cases} 88.5 \le 3x_1 + 3.5x_2 + \cdots + 6.5x_8 \le 88.9, \\ 19 \le \sum\limits_{i=1}^{8} x_i \le 20, \\ x_i \in N, \quad i = 1, 2, \cdots, 8. \end{cases}$$

通过Matlab[1]编程得出该模型的解A_i有90849个，即构成一捆第I类成品的方案有90849个．类似分类模型(一)，为了得到构成一捆第II类、第III类成品的方案，我们建立分类模型(二)和分类模型(三)如下：

分类模型(二)

$$\begin{cases} 88.5 \le 7x_1 + 7.5x_2 + \cdots + 13.5x_8 \le 89.5, \\ 7 \le \sum\limits_{i=1}^{14} x_i \le 8, \\ x_i \in N, \quad i = 1, 2, \cdots, 14. \end{cases}$$

分类模型(三)

$$\begin{cases} 88.5 \le 14x_1 + 14.5x_2 + \cdots + 22.5x_24 \le 89.5, \\ 4 \le \sum\limits_{i=1}^{24} x_i \le 5, \\ x_i \in N, \quad i = 1, 2, \cdots, 24. \end{cases}$$

通过Matlab[1]编程(程序类似分类模型(一))，得出分类模型(二)的解B_i有10866个，分类模型(三)的解C_i有4286个．

4.2 每类成品的最大捆数模型

从4.1的结果可知，对于每类产品，构成一捆的方法最少的都达到4286，若对所有的构成方案均进行考虑，则运算量就会太大，不易在30分钟内得出最多捆数的结果，因此我们对这些构成方案进行筛选．将构成此方案所需的原料种类超过三个的均舍弃，例如在构成一捆第I类成品的方案$3\times1+3.5\times4+4\times3+4.5\times4+5\times6+5.5\times1+6\times1=88.5$，此方案包含了3–3.4、3.5–3.9、4–4.4、4.5–4.9、5–5.4、5.5–5.9、6–6.4共七种原料，故将此方案舍弃掉．此假设在实际加工中也是合理的，因为构成一捆肠衣的方案所需的长度范围越多其工作量就越大，利用此假设可降低实际的工作量．

基于上述假设，我们对4.1中得出的结果进行筛选，得出构成第I类、第II类、第III类成品的方案分别有668、439、792种，具体结果见表1、表2和表3．

我们先忽略第四个条件，即在原料不可降级使用条件下的最大捆数．由第一个假设可知要使总捆数最大，实际上是三类成品在现有原料的条件下各自最大捆数的和，故我们先分别对三类模型求出最大捆数．下面我们以第I类成品为例建立模型，第II类、第III类产品则类似．

筛选后能够构成一捆第I类产品的方案有668种，而该成品的原料限制，如表1．

现假设第i种构成一捆的方案$A_i=(a_{i1},a_{i2},\cdots,a_{i8})$ $(i=1,2,\cdots,668)$有l_i捆，各长度范围根数为$o=(o_1,o_2,\cdots,o_8)=(43,59,39,41,27,28,34,21)$，则可知捆数$l_i$需满足各长度范围的

根数总数应少于o_i，如长度范围位于3–3.4的应满足$\sum_{i=1}^{668} a_{i1}x_i \le o_1$．由此可知，要在满足原料的条件下捆数最大实际上是一个求最大的整数规划问题[2-4]，现建立第I类成品的最大捆数模型如下

$$\max \quad \sum_{i=1}^{668} l_i$$

$$\text{s.t.} \quad \sum_{i=1}^{668} l_i a_{ij} \leqslant o_j, \quad j=1,2,\cdots,8,$$

$$l_i \in N, \quad i=1,2,\cdots,668.$$

该模型可以通过用Lingo编程[5]求解，该程序的运行结果见表2．表2中非0变量x_i所对应的方案A_i的具体数值，见表3．

表1: 第I类产品的原料

长度	3–3.4	3.5–3.9	4–4.4	4.5–4.9	5–5.4	5.5–5.9	6–6.4	36.5–6.9
根数	43	59	39	41	27	28	34	21

表2: 第I类成品的最大捆数模型的解

变量	x_7	x_{131}	x_{227}	x_{294}	x_{431}	x_{462}	x_{473}	x_{514}	x_{535}	x_{615}	其他
结果	2	1	1	1	1	2	1	1	1	3	0

表3: 第I类成品的最大捆数模型中非0解所对应的形成一捆产品的具体方案

方案	3–3.4	3.5–3.9	4–4.4	4.5–4.9	5–5.4	5.5–5.9	6–6.4	36.5–6.9
A_7	0	0	4	15	1	0	0	0
A_{131}	3	0	0	1	15	0	0	0
A_{227}	0	4	9	0	0	7	0	0
A_{294}	7	2	0	0	0	11	0	0
A_{431}	6	0	0	0	7	0	6	0
A_{462}	9	0	0	3	0	0	8	0
A_{473}	1	9	0	0	0	0	9	0
A_{514}	0	0	16	0	0	0	3	1
A_{535}	7	0	0	0	0	10	0	2
A_{615}	0	13	1	0	0	0	0	6

由表2和表3的结果可知第I类成品的最大捆数为14捆，此时原料的应用情况见表4.

表4: 第I类成品的最大捆数模型中原料的应用情况与给出的原料数量的比较

长度	3–3.4	3.5–3.9	4–4.4	4.5–4.9	5–5.4	5.5–5.9	6–6.4	36.5–6.9
给出原料根数	43	59	39	41	27	28	34	21
使用根数	42	54	36	37	24	28	34	21

由表4可知，第I类成品的最大捆数模型中对于给出的原料最终只剩下16根，不能构成一捆，因此第I类成品的最大捆数为14捆是合理的. 类似第I类成品的最大捆数模型，对于第II类、第III类产品可建立如下模型. 第II类成品的最大捆数模型：

$$\max \quad \sum_{i=1}^{439} l_i$$

$$\text{s.t.} \quad \sum_{i=1}^{439} l_i b_{ij} \leqslant p_j, \quad j=1,2,\cdots,14,$$

$$l_i \in N, \quad i=1,2,\cdots,439.$$

第III类成品的最大捆数模型：

$$\max \quad \sum_{i=1}^{792} l_i$$

$$\text{s.t.} \quad \sum_{i=1}^{792} l_i c_{ij} \leqslant q_i, \quad j=1,2,\cdots,24,$$

$$l_i \in N, \quad i=1,2,\cdots,792,$$

其中$p=(p_1,p_2,\cdots,p_{14})=(24,24,20,25,21,23,21,18,31,23,22,59,18,25)$为第II类成品的原料数，$q=(q_1,q_2,\cdots,q_{24})=(35,29,30,42,28,42,45,49,50,64,52,63,49,35,27,16,12,2,0,6,0,0,0,1)$为第III类成品的原料数.

对于第II类、第III类成品的最大捆数模型的求解算法类似于第I类成品的最大捆数模型，得到第II类成品最大捆数为37，对于给出的原料剩余58根. 由于该模型求出的所有结果构成一捆第II类成品的方案中均未用到长度为7.5–7.9的原料，故该长度范围的原料全部剩余，而长度为7–7.4的也只用了一根，因此造成了第II类成品的原料剩余量较多. 虽然如此，但长度位于8.5–13.4的原料完全用完，既长度较长的原料全部用完，这样有利于减少浪费，故第II类成品最大捆数模型也具有一定的合理性.

第III类成品得到的最大捆数为137，恰好是给出原料的根数. 所有的原料被完全用完，很显然第II类成品最大捆数模型是最优的，各方案安排也是最合理的.

综合第I类、第II类、第III类成品的最大捆数模型，可知在不考虑降级情况下总捆数最大为188捆. 虽然此时第III类成品的原料全部用完，但第I类、第II类的原料还有部分剩余，故模型有待改进.

5 模型的改进—综合最大捆数模型

第4部分给出了不考虑降级情况下的第I类、第II类、第III类成品最大捆数模型，得出最大捆数为188捆．但在这些模型中有很多材料被浪费掉，特别是长度为7.5–7.9的肠衣完全没有用到，7–7.4的肠衣虽然用到但也只用了一根，这对于大批生产的工厂来说损耗是很大的．为了减少材料的浪费，我们考虑将上一规格产品的材料剩余量降级使用．由第二个假设可知，原料降级使用时降到距离所在规格最近的长度范围，即第III类成品的剩余原料可以与第II成品的原料进行捆扎，降级为13.5–13.9的原料使用．第II类成品的原料与第I类成品的原料进行捆扎，降级为6.5–6.9的原料使用．基于以上分析，我们可以求解第III类成品最大捆数模型，得出结果后看是否有剩余的原料，若无则正好分配完，继续解第II类成品最大捆数模型，若有剩余，则将剩余的原料作为13.5–13.9的原料，再求解第II类成品最大捆数模型，以此类推，直至求完第I类成品最大捆数模型，最后将三个模型的结果求和就得到最大捆数．此方法虽容易理解，但过程繁琐，因此我们需建立一个统一的模型来求出最大捆数．用$x_i\,(i=1,2,\cdots,668)$表示选择第i种构成一捆第I类成品的方案$A_i=(a_{i1},a_{i2},\cdots,a_{i8})$的捆数，$y_i\,(i=1,2,\cdots,439)$表示选择第$i$种构成一捆第II类成品的方案$B_i=(b_{i1},b_{i2},\cdots,b_{i14})$的捆数，$z_i\,(i=1,2,\cdots,792)$表示选择第$i$种构成一捆第III类成品的方案$C_i=(c_{i1},c_{i2},\cdots,c_{i24})$的捆数．第I类产品有8种不同长度的原料3–3.4表示第1种原料，3.5–3.9为第2种原料，以此类推至第8种原料．类似地，第II类、第III类产品的原料分别为14、24种．第I类产品的第i种原料在最优分配后的剩余量用$d1_i\,(i=1,2,\cdots,8)$表示，第II类产品的第i种原料在最优分配后的剩余量用$d2_i\,(i=1,2,\cdots,14)$表示，第III类的第i种原料在最优分配后的剩余量用$d3_i\,(i=1,2,\cdots,24)$表示，则考虑原料可以降级使用下的综合最大捆数模型为：

$$
\begin{aligned}
\max\quad & \sum_{i=1}^{668} x_i+\sum_{i=1}^{439} y_i+\sum_{i=1}^{792} z_i \\
\text{s.t.}\quad & \sum_{i=1}^{668} a_{ij}x_i+d1_j=o_j,\quad j=1,2,\cdots,7, \\
& \sum_{i=1}^{668} a_{i,8}x_i+d1_8-\sum_{i=1}^{1} 4d2_i=o_8, \\
& \sum_{i=1}^{439} b_{ij}y_i+d2_j=p_j,\quad j=1,2,\cdots,13, \\
& \sum_{i=1}^{439} b_{i,14}y_i+d2_{14}-\sum_{i=1}^{24} d3_i=p_{14}, \\
& \sum_{i=1}^{792} c_{ij}z_i+d3_j=q_j,\quad j=1,2,\cdots,24, \\
& x_i\in N,\quad i=1,2,\cdots,668,\quad y_i\in N,\quad i=1,2,\cdots,439, \\
& z_i\in N,\quad i=1,2,\cdots,792,\quad d1_i\in N,\quad i=1,2,\cdots,8, \\
& d2_i\in N,\quad i=1,2,\cdots,14,\quad d3_i\in N,\quad i=1,2,\cdots,24.
\end{aligned}
$$

在此模型中，$d1_i$, $d2_i$ 与 $d3_i$ 相当于线性规划中的松弛变量或人工变量．此模型是典型的整数规划模型，可以通过Lingo软件求解．

在第4部分我们已经得出在最大捆数情况下第III类产品的原料正好全部用完，而第I类产品的原料不存在降级使用的情况，为了减少模型的变量，减少求解难度，我们将上述综合最大捆数模型进行简化，得到简化后的综合最大捆数模型如下：

$$\max \quad \sum_{i=1}^{668} x_i + \sum_{i=1}^{439} y_i + \sum_{i=1}^{792} z_i$$

$$\text{s.t.} \quad \sum_{i=1}^{668} a_{ij} x_i \le o_j, \quad j = 1, 2, \cdots, 7,$$

$$\sum_{i=1}^{668} a_{i,8} x_i - \sum_{i=1}^{1} 4d2_i = o_8,$$

$$\sum_{i=1}^{439} b_{ij} y_i + d2_j = p_j, \quad j = 1, 2, \cdots, 13,$$

$$\sum_{i=1}^{439} b_{i,14} y_i + d2_{14} = p_{14},$$

$$\sum_{i=1}^{792} c_{ij} z_i = q_j, \quad j = 1, 2, \cdots, 24,$$

$$x_i \in N, \quad i = 1, 2, \cdots, 668, \quad y_i \in N, \quad i = 1, 2, \cdots, 439,$$

$$z_i \in N, \quad i = 1, 2, \cdots, 792, \quad d2_i \in N, \quad i = 1, 2, \cdots, 14.$$

显然该简化模型比原来的综合最大捆数模型减少了 $d1_i$ 与 $d3_i$，我们通过Lingo编程求解该程序，结果如表5所示，而综合最大捆数模型中原料的应用情况与给出的原料数量的比较分析，结果如表6所示．

表5: 综合最大捆数模型的解

变量	x_{148}	x_{315}	x_{486}	x_{525}	x_{572}	x_{604}	x_{633}	x_{667}	y_{10}	y_{34}	y_{43}	y_{79}	y_{149}	y_{166}	y_{169}
结果	1	2	1	2	1	2	4	2	3	3	4	1	2	1	5
变量	y_{177}	y_{204}	y_{206}	y_{225}	y_{258}	y_{338}	y_{350}	y_{367}	z_2	z_6	z_{13}	z_{15}	z_{30}	z_{34}	z_{48}
结果	1	1	3	2	2	2	3	4	7	3	1	1	2	2	1
变量	z_{49}	z_{58}	z_{67}	z_{79}	z_{89}	z_{96}	z_{105}	z_{107}	z_{110}	z_{113}	z_{120}	z_{125}	z_{129}	z_{132}	z_{144}
结果	10	13	1	2	5	3	1	7	10	1	1	4	1	2	3
变量	z_{149}	z_{154}	z_{157}	z_{183}	z_{184}	z_{189}	z_{207}	z_{209}	z_{210}	z_{217}	z_{227}	z_{231}	z_{234}	z_{235}	z_{240}
结果	1	1	1	11	4	10	2	3	1	2	1	1	2	4	21
变量	z_{258}	z_{268}	z_{270}	z_{344}	z_{431}	z_{717}	其他								
结果	1	1	1	2	6	1	0								

表6：综合最大捆数模型中原料的应用情况与给出的原料数量的比较

长度	3–3.4	3.5–3.9	4–4.4	4.5–4.9	5–5.4	5.5–5.9	6–6.4	6.5–6.9
给出根数	43	59	39	41	27	28	34	21
使用根数	42	59	39	40	23	27	33	21
长度	7–7.4	7.5–7.9	8–8.4	8.5–8.9	9–9.4	9.5–9.9	10–10.4	10.5–10.9
给出根数	24	24	20	25	21	23	21	18
使用根数	15	24	10	25	21	23	21	18
长度	11–11.4	11.5–11.9	12–12.4	12.5–12.9	13–13.4	13.5–13.9	14–14.4	14.5–14.9
给出根数	31	23	22	59	18	25	35	29
使用根数	31	23	22	5	18	25	35	29
长度	15–15.4	15.5–15.9	16–16.4	16.5–16.9	17–17.4	17.5–17.9	18–18.4	18.5–18.9
给出根数	30	42	28	42	45	49	50	64
使用根数	30	42	28	42	45	49	50	64
长度	19–19.4	19.5–19.9	20–20.4	20.5–20.9	21–21.4	21.5–21.9	22–22.4	22.5–22.9
给出根数	52	63	49	35	27	16	12	2
使用根数	52	63	49	35	27	16	12	2
长度	23–23.4	23.5–23.9	24–24.4	24.5–24.9	25–25.4	25.5–25.9		
给出根数	0	6	0	0	0	1		
使用根数	0	6	0	0	0	1		

由表5可知，综合最大捆数模型所得的最大捆数为

$$\sum_{i=1}^{668} x_i + \sum_{i=1}^{439} y_i + \sum_{i=1}^{792} z_i = 191,$$

比分类最大捆数模型的总和要多3捆．由于篇幅有限，对于在情况最优下，各个 x_i, y_i, z_i 所对应的具体构成一捆产品的方案在此不再列出，在表1、表2与表3中可查得相应的值．在综合最大捆数模型下得到第I类、第II类、第III类成品的捆数分别为17、37、137捆．我们会发现第I类成品的捆数比我们给出的材料所能构成的捆数要多，这就是原料降级使用的结果，因为我们的第II类成品原料有剩余，将其原料降为长度为6.5–6.9的原料使用，最终使得我们的第I类成品比分类最大捆数模型中多出3捆．实际上在表6中长度为7–7.4与7.5–7.9的原料的使用根数均是降为6.5–6.9所使用的根数，在本来的第II类成品最大中均未使用．

6　模型的总结与评价

本文在合理假设的基础上，首先通过建模得到构成一捆产品的全部方案．由于所得方案的总量过大，对于30分钟内得出分配方案无法做到，因此我们对方案进行筛选，舍弃跨越三个原料种类以上的方案，得出构成一捆第I类、第II类、第III类成品的方案分别有668、439、792种．对于筛选后得到的方案相比最初的数据减少了很多，而且在目前高科技飞速发展的时代，对于这个数据在30分钟内得出方案完全有可能．当然也可以在此基础上对方案再进行限制，得到更少的方案，但是这样有可能使某些长度范围的原料在任何一种方案中

都未使用到，那么此时对于那些未使用到的原料就非常的浪费．而在我们筛选后得出的方案中未出现这种状况，故筛选是合理的．在得出构成一捆成品的方案后，我们先建立在原料不可降级使用时的捆数最大模型．由原料不可分割的假设可知要想总捆数最大，实际上是三类成品在现有原料的条件下各自最大捆数的和，所以我们建立了分别对应三类产品的最大捆数模型，实际上也是含不等式约束的线性规划问题．第I类、第II类、第III类产品最大捆数模型下，我们得到它们的最大捆数分别为14捆、37捆、137捆．虽然第III类产品的原料全部用完，但第I类、第II类产品的原料还存在部分原料的剩余．三类产品分成三个模型，虽然能控制在30分钟内，但非常繁琐，而且没有考虑原料降级使用的条件，对原料的浪费较多，存在一些不合理的地方．由于分类最大捆数模型存在一系列的不足，所以我们又在考虑原料可降级使用的条件下，得出一个全部产品一起考虑的模型，即综合最大捆数模型．该模型在之前分类最大捆数模型的基础上，加入一些松弛变量或人工变量，并对三类产品最大捆数模型进行综合，得出一个综合整数规划模型，求解得到最大捆数为191捆，比分类时多3捆．该模型达到最优解时，第III类产品的原料全部用完，故满足了最短长度最长的产品最多的条件．综合最大捆数模型既考虑了原料的降级使用，也能通过一个模型将总的最大捆数得出．虽然该模型的变量较多，但通过Lingo来求解整数规划模型速度很快，通过我们给出的程序，几分钟就得出了最终结果．因此对于一般情况的原料，在已得出的构成一捆产品方案的基础上，再应用我们给出的Lingo程序，在30分钟之内一定可以得出最大捆数方案．虽然最终的模型考虑到了所有的因素，但对于降级的问题我们只是将其降为下一等级长度范围最长的情况，对于是否可以降为其它长度未考虑，因此有待进一步的改进．

参考文献：

[1] 胡良剑，孙晓君．Matlab数学实验[M]．北京：高等教育出版社，2006

Hu L J, Sun X J. Matlab Mathematical Experiment[M]. Beijing: Higher Education Press, 2006

[2] 甘应爱，等．运筹学(第3版)[M]．北京：清华大学出版社，2005

Gan Y A, *et al.* Operations Research (3rd Edition)[M]. Beijing: Tsinghua University Press, 2005

[3] 彭放，杨瑞瑛，等．数学建模方法[M]．北京：科学出版社，2007

Peng F, Yang R Y, *et al.* Mathematical Modeling Method[M]. Beijing: Science Press, 2007

[4] 韩中庚．数学建模方法及其应用[M]．北京：高等教育出版社，2005

Han Z G. Mathematical Modeling Method and its Application[M]. Beijing: Higher Education Press, 2005

[5] 谢金星，薛毅．优化建模与Lindo\Lingo软件[M]．北京：清华大学出版社，2005

Xie J X, Xue Y. Optimization Modeling and Lindo\Lingo Software[M]. Beijing: Tsinghua University Press, 2005

Study on the Maximum Number of Bundles about Collocation of Natural Casing

ZHOU Ya, LIU Zhen-tang, LU Xue-yang

Advisor: HUANG De

(Guangxi City College, Guangxi 532100)

Abstract: This paper attains all product schemes of constituting a bundle of natural casing from the company's specific requirements. Because the scale of the attained schemes is too big, we abandon those raw materials more than three kinds of above schemes, and build the bundle number maximum model in the presumption that the raw material cannot be relegated to use. In order to improve the utilization rate of the raw materials, this paper considers that the raw materials can be degraded to use, and draw a general integer programming model for all the products. The optimal solution is calculated from all of the model solution bundles. The optimal solution is 191 bundles, and thus the optimum matching scheme can be determined.

Keywords: integer programming; classification scheme; Matlab; Lingo